desk

Vol. 46 APRIL, 1955 No. 4

# The Journal of Educational Psychology

UNIVERSITY OF MICHIGAN
JUN
PERIODICAL READING ROOM

## CONTENTS

*Technique in Problem-solving as a Predictor of Educational Achievement* ..... 193
K. PATRICIA CROSS AND EUGENE L. GAIER

*Normative Test Data for Adult Males Obtained by House-to-house Testing* ..... 207
ELIZABETH P. HAGEN AND ROBERT L. THORNDIKE

*Predicting Attrition-survival in First Year Engineering* ..... 217
JOHN P. MALLOY, BOLESLAW WYSOCKI AND LEO F. GRAHAM

*Bilingualism and Biculturalism* ..... 222
JAMES P. SOFFIETTI

*War-time Accelerates Ten Years After* ..... 228
MARIE A. FLESHER AND S. L. PRESSEY

*A Note on the Standard Error of Measurement* ..... 239
C. H. PATTERSON

*The Stability of Interest Scores of College Freshmen* ..... 243
CLAY E. GEORGE AND ALBERT J. KINGSTON

*Preliminary Exploration of Personality Factors Related to Success in a High-level Military School* ..... 247
JAMES H. L. ROACH AND EDWARD M. WALL

Book Reviews ..... 251

Published monthly except June to September
$7.00 a year in U. S. and Pan Amer.; Canada, $7.20; other countries, $7.40.
Single issues $1.20

WARWICK & YORK, INC.
BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879.

# THE JOURNAL OF Educational Psychology

*Established 1910*


BOARD OF EDITORS*

Wm. Clark Trow, *Chairman*
University of Michigan

Stephen M. Corey
Teachers College
Columbia University

J. B. Stroud
State University of Iowa

* Appointed by the American Psychological Association.

EDITORIAL CONSULTANTS

Paul Blommers
Chester W. Harris
Harold E. Jones
H. H. Remmers
Percival M. Symonds
Miles A. Tinker
Alexander G. Wesman
Paul A. Witty


The Journal of Educational Psychology is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The Journal will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

## Information for Contributors

Manuscripts and communications regarding editorial matters should be addressed to the Chairman of the Board of Editors, Wm. Clark Trow, School of Education, University of Michigan, Ann Arbor, Mich.

The Journal has set regulations regarding content and style of material published, and these should be observed in the preparation of manuscripts to be submitted.

*Tables and graphs.*—In keeping with the practice of most professional periodicals of like character, The Journal now requires authors to bear part of the added cost resulting from the use of tables, formulas, and graphs. Prospective contributors may obtain detailed statement as to

*(Continued on Inside Back Cover)*

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 April, 1955 Number 4

# TECHNIQUE IN PROBLEM-SOLVING AS A PREDICTOR OF EDUCATIONAL ACHIEVEMENT[1, 2]

K. PATRICIA CROSS AND EUGENE L. GAIER

University of Illinois

Successful problem-solving involves both the knowledge of fundamental facts and their effective utilization in arriving at a solution (*1*, *3*). Familiar to teachers and employers is the student or worker who pays little attention to the interrelationships among the facts used. Presumably, he is limited to previously encountered specific situations for which he has memorized the solutions. At the other end, is the individual who has concentrated on an 'understanding' of the general principles underlying the problem. He is not dependent upon rote recall but comprehends and employs the principles involved. No inherent conflict exists, however, between the learning of content and the development of analytical acumen. These processes can be considered not as opposed but as complementary goals of education.

This paper reports the development of a test—the Balance Problems Test (BPT)—designed to assess individual differences in problem-solving. The test affords indices of relative preference for the use of facts versus the use of principles as a technique in problem-solving. The following hypotheses are proposed:

1) Proficiency in solving problems on the test will be positively related to the tendency to use principles rather than facts alone.

[1] This research was supported in part by the United States Air Force under Contract No. 33(038)-25726 monitored by Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] The writers wish to express appreciation to Dr. L. L. McQuitty for his critical judgments and stimulating interest in this study.

2) Both the number of problems solved and the indices of relative dependency upon principles over facts in the BPT will be positively related to measures of educational achievement.

3) Scores on the BPT, reflecting degrees of preference for 'rational' over 'routine' techniques in problem-solving, will compare favorably with tests of 'general aptitude' as predictors of 'problem-solving achievement' (e.g., as indicated by achievement in mathematics).

## THE BALANCE PROBLEMS TEST

In the BPT, problems may be solved either through applications of facts[3] for each specific problem or through the use of a general principle applicable to an entire set of problems. The use of principles is assumed to be a more difficult approach than the use of facts, and a slight reward is offered for the use of this method. However, should the subject be either unable or unwilling to put forth the effort required for solution by principles, he is free to use facts. It was thought that a test of this kind would provide a better prediction of achievement in problem-solving through inclusion of an index of preference of method as well as a measure of ability.

*Structure of the Test.*—(Test design) The Balance Problems Test consists of six sets of problems, each set containing four problems. Each set of problems may be solved by two different methods: (a) use of a single principle, or (b) use of four facts. (The subject may, of course, use some combination of these two methods.) The general principle applies to the entire set of four problems whereas each fact is applicable to only one problem. The solution by facts requires less effort and less time than the solution by a principle. Consequently, it was deemed desirable to word the instructions so that a somewhat higher value was placed on the use of the principle. The subject was to be free, however, to select either method, or a combination of both methods, as he might desire. An incentive for using the principle was provided by informing the subject that

[3] By 'facts' we mean information which might be utilized in solving a specific problem or a limited category of problems, requiring but simple recall and direct application. By 'principles', we mean information which can be employed in solving a larger class or variety of problems. The use of principles in solving problems requires more complex thought, such as the perception of relationships, the eduction of correlates, as well as a knowledge of the ideas contained in the problem.

one point would be charged for each piece of 'information' used in solving the problem. For each set of four problems, the use of facts carried a penalty of four points whereas only one penalty point would be charged for use of the principle. The penalties were not used directly in scoring the tests, although one of the scores does place a penalty on the use of facts. The primary function of these instructions was to encourage the use of the principle.[4]

(Test format) Each page of the test contains one set of four problems and the information necessary for its solution (hidden from view), as shown in Figure 1.

All sets of problems have the general format shown in Figure 1, but the figures and complexity of the problems change with each set. The sets become increasingly difficult as the subject proceeds through the test.

On the left-hand side of the page, four problems were presented. On the right-hand side of each page of the test is the necessary information for the solution of the problems of a set. The 'Principle' which is applicable to each set is hidden under a tab at the top; beneath are the four facts, all hidden under tabs which must be removed before the information can be read. Once the tab has been removed, it is irreplaceable; the scorers are able to count the number of information tabs seen by the subject. It should be noticed, however, that this does not tell what information was actually employed or how the information was utilized for the actual solution. Many instances were observed in which an effort was first made to use the principle; then, upon failing to solve the problem, facts became the information utilized.

After the tabs had been removed, the principles and facts for the given problem set would appear as shown in Figure 1.

*Scoring the Test.*—Five scores on the test were derived:

(Variable 1: Number correct.) This score is a count of the number of problems answered correctly irrespective of the kind of in-

---

[4] It is recognized at the outset that these instructions may have seemed to make this test dependent, in part, on individual differences in the effectiveness of the incentive. But, since the incentive was objectively the same for all subjects, individual differences in its effectiveness are an index of the relative strengths of preferences for facts versus principles. Abstract rather than general mechanical principles were employed here so that a minimum of specific knowledge would be required and differences in degrees of training would be minimized.

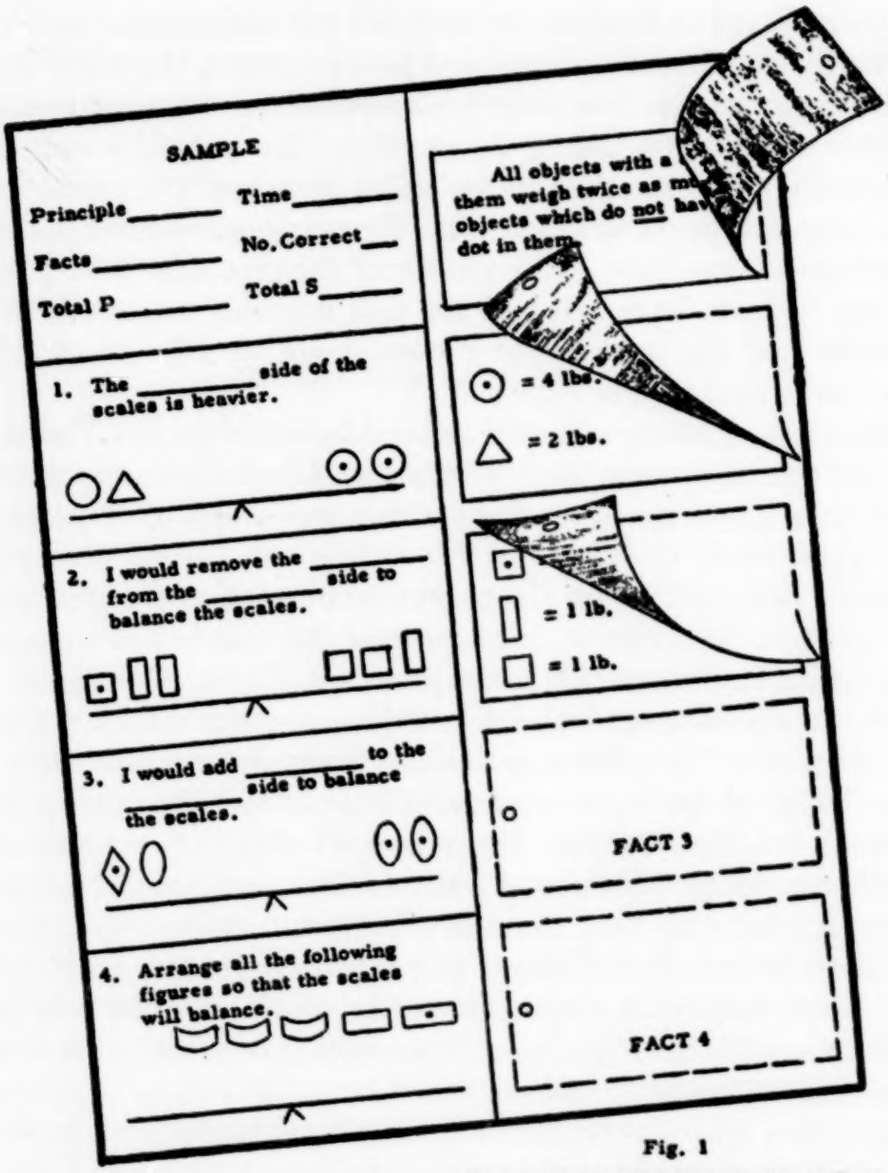
SAMPLE
Principle_____ Time_____
Facts_____ No. Correct___
Total P_____ Total S_____
1. The _____ side of the scales is heavier.
2. I would remove the _____ from the _____ side to balance the scales.
3. I would add _____ to the _____ side to balance the scales.
4. Arrange all the following figures so that the scales will balance.
= 4 lbs.
= 2 lbs.
= 1 lb.
= 1 lb.
FACT 3
FACT 4

**Fig. 1**

**Sample Page of Balance Problems Test**

FIG. 1, "Techniques in problem-solving" by Cross and Gaier, pp. 193–206 Apr. Jour. Ed. Psy.

formation used by the subject in his solution. This score was unweighted, the possible range being from zero to a maximum of twenty-four correct.

(Variable 2: Number of facts uncovered.) This score is a frequency count of the number of facts uncovered during the test. No attempt was made here to evaluate the effectiveness of the subject's use of the information uncovered; it is merely the number of 'Fact' tabs torn off, the information being exposed and pre-

sumably read. The possible score ranges from zero to a maximum of twenty-four facts uncovered.

(Variable 3: Number of principles uncovered.) This score is a frequency count of the number of principles uncovered during the test. As in Variable 2, no attempt was made to evaluate the subject's effective application of the information uncovered. Rather, this score is merely the number of 'Principle' tabs torn off. The possible score ranges from zero to a maximum of six.

(Variable 4: Time.) Number of minutes used in working the test.

(Variable 5: Weighted test score ($S = C + 24 - F$).) The weighted score is the number correct + 24 − number of facts uncovered. The constant of twenty-four was inserted to eliminate negative scores in those cases where number of facts uncovered exceeded the number of correct answers. The weighted score was developed as a joint measure of the influence of facts and principles; the fewer facts a subject uncovered, the more he was dependent on principles. The subject who correctly solved all the problems through the use of facts received only half as large a score as the subject who obtained all the correct answers through the use of principles exclusively.

Of the five variables described above, it was thought that the weighted score would be the best predictor of high school mathematics ability. Achievement in mathematics is often presumed to be related to individual differences in the somewhat generalized effective use of principles in solving problems; the weighted score represents an effort to evaluate these differences.

*Pre-testing the Instrument.*—A preliminary administration of the test was conducted for the purposes of determining approximate testing time, difficulty level, and the comprehensibility of the instructions. For these purposes, the test was administered individually in unspeeded sessions to five high school students and to five airmen who were experienced airplane and engine mechanics.

The test was next given to fourteen experienced airplane and engine mechanics at Chanute Air Force Base, in two group sessions of seven subjects each. This testing resulted in further clarification of instructions and in the revision of some of the problems to a higher level of difficulty. The subjects were carefully observed as they began solving the test problems by the use of principles, as well as when they shifted from principles to the use of facts in their attempt to solve the problems. Upon completion of this pre-

test, the test was further revised into the form described above, for use in the present study.

## PROCEDURE

Our second hypothesis maintains that relationships exist between BPT scores and student achievement. In order to test this hypothesis, it was necessary to administer the test to a group of subjects for whom appropriate criterion data could be secured. It seemed that such data should include diverse indices of educational achievement for subjects reasonably homogeneous in general social background and educational level. It was possible to secure five 'criterion scores' for the subjects described below; these scores are called 'Variables 6, 7, 8, 9, and 10', with reference to their representation in the correlational analysis.

In addition to the five BPT scores and the five indices of educational achievement, two measures of general aptitude (Language and Non-language Scores) were available for all subjects.

*Subjects.*—Twenty-five juniors and fourteen seniors (one entire study class) in a high school of eight hundred students served as subjects. The school is located in an Illinois city of approximately 15,000 inhabitants. In addition to BPT scores, the following criterion and prediction data were gathered for this sample:

*Criteria.*—The selection of criteria of student achievement was governed by considerations indigenous to the test, as well as those unique to the educational background and training of high school subjects. Considerable evidence has been reported concerning the relationship of abstract and relational thinking to mathematical aptitude and training. Since the use of principles in the BPT demanded the ability to generalize a given relationship to a set of problems, it appeared desirable to evaluate the BPT in relation to indices of actual mathematical training and achievement. Thus, a BPT score was to be employed as one of the predictors of the students' previous years' mathematics achievement (i.e., average mathematics grades), and compared with other predictors.

(Variable 6: Mathematics level.) This variable is essentially a keyed score for estimating the highest level of mathematical training reached. The subjects were asked to check on the cover page of the test those mathematics courses they had taken. The list presented contained the following courses: (a) Elementary Algebra; (b) Plane Geometry; (c) Solid Geometry; (d) Advanced Algebra;

(e) Trigonometry; and (f) any College Mathematics. In addition, the subject was asked to list any other mathematics courses studied. It was noticed that the categories and number of mathematics courses taken frequently seemed to represent different levels of mathematics sophistication. For example, it appeared that a differentiation should be made in the mathematics sophistication of the students of general or commercial mathematics and elementary algebra. Likewise, the person taking college mathematics should have reached a level of mathematical functioning not necessarily dependent upon how much high school mathematics he had had. For these reasons, it was concluded that mathematics level, wherein each course was differentially weighted,[5] would be a more sensitive measure of mathematical sophistication than would the actual number of mathematical courses taken. As a check on this, the degree of correlation was determined between mathematical level and number of mathematics courses completed. A product-moment correlation value of 0.98 was obtained. In the subsequent analyses, the mathematics level has been used.

(Variable 7: Average mathematics grade.) The average mathematics grade consisted of averaging the numerical grades for all courses taken. The average was based on anywhere from zero to six grades, depending upon the number of mathematics courses taken. It would have been interesting to have had a measure of the achievement of all subjects at any one level; but, the subjects were never all of the same level in the courses they were taking.

(Variable 8: Yearly Average Grade.) The yearly average grade was obtained by averaging the numerical grades in all the courses taken by the students during the previous year. The number of courses taken by the students ranged from four to six, inclusively. Since the mathematics course grade was included for those who took mathematics, the yearly average grade, in many instances, was not independent of the Average Mathematics Grade (Variable 7).

(Variable 9: Physical Science Achievement.) An index of ability

---

[5] Each course was weighted as follows, and a subject's index of mathematical sophistication was a sum of the weights for the courses he took: 0 = no mathematics courses taken; 1 = each course in advanced arithmetic and commercial mathematics; 2 = elementary algebra; 3 = plane geometry; plus one additional point for each of the following courses—advanced algebra, solid geometry, trigonometry; 7 = any college mathematics course.

to read and comprehend material in the natural sciences (also, ability to read and interpret graphs and tables). This is a standardized test score obtained from the Illinois Test of Reading Comprehension.

(Variable 10: Social Science Achievement.) An index of ability to read and comprehend social science material. Like Variable 9, this score was obtained from the Illinois Test of Reading Comprehension.

*Other Predictors.*—Since the school from which our sample was obtained was one of the participating members of the Illinois Statewide High School Testing Program, standardized indices of aptitude as well as achievement were available.

Two scores of 'general aptitude' were selected for use as predictors, both obtained from the California Short-Form Test of Mental Maturity:

(Variable 11: Language.) An index of ability to manipulate verbal symbols.

(Variable 12: Non-language.) An index of ability to deal with abstract material, other than verbal.

## RESULTS

In order to test the hypotheses stated above, product-moment correlations based on the total sample of thirty-nine subjects were computed among all twelve variables utilized in this study. The obtained coefficients are presented in Table 1. It should be noted that for the number of cases used here, product-moment $r$'s must equal or exceed .27 to be significant at the five per cent level, and .37 to be significant at the one per cent level of confidence. Both of these values are based on a single-tailed test, i.e., where the direction of the $r$ is hypothesized, as is the case for all the hypotheses of this study.

The correlation coefficients of Table 1 will be considered in relation to testing each of the hypotheses outlined above; and, in relation to the possibility of yielding other hypotheses for subsequent investigations.

*First hypothesis.*—According to this hypothesis, proficiency in solving problems on the test is positively related to the tendency to use principles. The measure of proficiency here used is the number of problems of the BPT correctly solved, Variable 1 of the table. The measures of the tendency to use principles are Variable

TABLE 1.—PRODUCT-MOMENT CORRELATION COEFFICIENTS FOR SCORES ON THE BALANCE PROBLEMS TEST AND SCHOOL ACHIEVEMENT DATA*

| Variables | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. No. correct | | | | | | | | | | | | |
| 2. No. facts uncovered | −29 | | | | | | | | | | | |
| 3. No. principles uncovered | 29 | −76 | | | | | | | | | | |
| 4. Time-minutes | 17 | −02 | 20 | | | | | | | | | |
| 5. C + 24 − F | 63 | −93 | 73 | 06 | | | | | | | | |
| 6. Math. level | 54 | −61 | 45 | −03 | 71 | | | | | | | |
| 7. Aver. math. grade | 28 | −46 | 34 | −03 | 49 | 33 | | | | | | |
| 8. Yearly aver. grade | 28 | −44 | 29 | −05 | 47 | 42 | 83 | | | | | |
| 9. Phys. Sci. Ach. | 61 | −37 | 24 | −03 | 54 | 57 | 52 | 45 | | | | |
| 10. Soc. Sci. Ach. | 32 | −60 | 36 | 02 | 61 | 34 | 52 | 55 | 66 | | | |
| 11. Language score | 42 | −45 | 35 | −18 | 53 | 60 | 55 | 63 | 71 | 66 | | |
| 12. Non-lang. score | 55 | −43 | 29 | −06 | 57 | 44 | 44 | 34 | 60 | 38 | 52 | |

* Decimal points have been omitted.

2, number of facts uncovered (an inverted index), and Variable 3, the number of principles uncovered. Each of these variables has a correlation of .29 with number of problems correctly solved. The correlation with the number of facts is negative and the one with number of principles is positive. These two $r$'s are only slightly more significant than the five per cent level, using a one-tailed distribution. These findings barely substantiate the hypothesis that proficiency on the test is positively correlated with the tendency to use principles. They indicate that subjects tend to be more proficient in problem-solving to the extent that they prefer to use principles rather than facts.

*Second hypothesis.*—According to the second hypothesis, both the number of problems solved and the indices of relative dependency upon principles over facts in the BPT are positively related to measures of educational achievement.

In Table 1, the number of problems correctly solved is designated by Variable 1. The tendency to use principles is assessed by Variable 2, number of facts uncovered (an inverted index) and Variable 3, number of principles uncovered. In addition, Variable 5 is a weighted combination of the number of problems correctly solved and the relative dependency on principles. Because it assesses jointly proficiency in problem-solving and the preference for prin-

TABLE 2.—PRODUCT-MOMENT CORRELATION COEFFICIENTS FOR SCORES ON BALANCE PROBLEMS TEST AND SUBJECT MATTER ACHIEVEMENT INDICES

| Variables | 1 | 2 | 3 | 5 |
|---|---|---|---|---|
| 6. Mathematics Level | .54** | −.61** | .45** | .71** |
| 7. Average mathematics grade | .28* | −.46** | .34* | .49** |
| 8. Yearly average grade | .28* | −.44** | .29* | .47** |
| 9. Physical Science Achievement | .61** | −.37** | .24 | .54** |
| 10. Social Science Achievement | .32* | −.60** | .36* | .61** |

* Significant at the five per cent level, but less than the one per cent level.

** Significant at beyond the one per cent level.

ciples, it was thought that it would be more highly related to achievement in mathematics than would either Variables 2 or 3.

The measures of achievement are repeated in Table 2 together with the *r*'s that they were found to have with the above mentioned variables of the BPT.

The only *r* not significant at the five per cent level or beyond is the value of .24 between Variables 3 and 9. The six correlations significant at the five per cent level or better, but less than the one per cent level, are indicated by an asterisk in Table 2. The remaining thirteen values are all significant at the one per cent level or beyond. These results support the second hypothesis that the number of problems solved and the indices of dependency on principles are positively related to indices of achievement.

The correlations of the weighted score (C + 24 − F) with indices of mathematical achievement are all higher than the analogous *r*'s with Variables, 1, 2, and 3. This gives support to the stated expectancy that the weighted index would be a better predictor of mathematics achievement than any of the other BPT indices.

The degree of association between the level of mathematics training and the weighted BPT score is .71. The BPT problems do not employ any of the so-called 'facts' taught in high school mathematics courses. Solution of the BPT problems does not depend upon the learning of specialized high school mathematical techniques. The application of a principle in the BPT requires no higher mathematics than does the use of facts; both can be applied without any mathematical training beyond elementary school arithmetic.

As shown in Table 2, the *r* for the weighted score (Variable 5) with an index of achievement is larger in every case except one than the analogous correlations using any one of the other measures on the BPT in lieu of the weighted score. The one exception is the correlation of .61 between Physical Science Achievement and number of BPT problems correctly solved. The analogous *r* with the weighted score is .54. In other words, a combined index of preference and proficiency in problem-solving is here indicated to be a better predictor of subject matter achievement than either of the indices separately taken. Since the two predictors, combined in this study on an *a priori* basis, have yielded promising results, it would appear worth while to pursue a much more elaborate study to determine the best relative weight to assign to each variable for predictive purposes.

The finding that Physical Science Achievement has lower correlations with the indices of the use of principles, Variables 2 and 3, than with indices of achievement in both mathematics and social science, (see Table 2) led us to examine the content of the tests of achievement. Observationally, the Physical Science test appeared to contain a greater proportion of items designed to test factual information than did the test of Social Science Achievement. The latter test contained more items requiring the use of principles for their solution. It also appears that mathematics grades in high school depend on effective use of principles. These tentative interpretations suggest the hypothesis that indices of the use of principles on the BPT are more highly related to achievement to the extent that the achievement depends more on effective use of principles rather than factual knowledge alone.

*Third hypothesis.*—In line with our third hypothesis, we were interested in the extent to which BPT scores would compare favorably with tests of general aptitude as predictors of problem-solving achievement, i.e., indicated by achievement in mathematics. Specifically, to what extent can the BPT contribute to the prediction of mathematical achievement?

To investigate this question, it was decided to determine what combination of test scoring variables and achievement scores would yield the best prediction of mathematics grades. The three measures selected to predict grades were weighted BPT scores (C + 24 − F), and Language and Non-language scores of the California Test of Mental Maturity. The intercorrelations with the criterion

(mathematical grades) are shown in Table 1. The obtained beta weights were .160, .393, and .216 for BPT (C + 24 − F), Language, and Non-language scores, respectively. The resulting regression equation was:

$$X_1' = .100\,X_2 + .080\,X_3 + .050\,X_4 + 71.74$$

where $X_1'$ = predicted mathematics grade

$X_2$ = weighted BPT score (C + 24 − F)
$X_3$ = Language Score
$X_4$ = Non-language Score

The multiple correlation obtained was:

$$R_{1.234} = .623$$

This multiple correlation can best be interpreted here in terms of $R^2$, the coefficient of multiple determination. This is the proportion of the variance in $X_1$ (average mathematical grade) that is dependent upon or predicted by $X_2$, $X_3$, and $X_4$ combined. In this case, $R^2$ was .3895, indicating that 38.95 per cent of the variance in average mathematics grades was accounted for by the weighted BPT score, the Language and Non-language scores, eliminating from multiple consideration whatever the scores have in common.

Of these three test scores, the results indicate that the BPT is the best single predictor of mathematics grades. At the same time, it should be recognized that the particular score chosen for the BPT (C + 24 −F) was the one which had yielded the highest correlation with average mathematics grades. This point is counteracted somewhat by the prediction made prior to initiating the testing of subjects: that this particular score, with its penalty for use of facts in solving the problems, would yield the best predictions (see description of Variable 5 above).

### SIGNIFICANCE OF THE BALANCE PROBLEMS TEST

The Balance Problems Test represents a unique contribution in predicting educational achievement. BPT indices of (a) proficiency in problem-solving, (b) degree of preference for principles over facts in problem-solving, and (c) a weighted combination of these are significantly correlated with achievement in three subject matter areas. The last index appears to be as promising, or even more

promising, than either a language or non-language test in the prediction of mathematics grades.

These conclusions have both psychological and educational significance, as well as practical ones. In addition to the predictive possibilities of the BPT, the results indicate that mode of intellectual attack on problems, as represented by individual differences in preference for principles, is a significant variable in problem-solving. If these findings continue to be substantiated, new educational and psychological research problems of determining the variables responsible for these individual differences will present themselves.

## SUMMARY AND CONCLUSIONS

The development of a new test—the Balance Problems Test—was undertaken in an attempt to measure the extent to which an individual selects and makes effective use of principles as opposed to factual information in problem-solving. By examining the kinds of information selected, it was hoped to tease out the individuals' characteristic problem-solving approach. Specifically, it was hypothesized that the individual who tended to use general principles would be superior to those depending upon the use of specific facts, in school achievement which involved problem-solving behavior.

Six sets of problems containing four sub-problems each were constructed for use in this study. The problems in each set could be solved either through the application of one general principle applying to the entire set of four problems or through the use of successive, discrete facts. All problems had exactly the same format, but the figures and complexity changed from problem to problem. Each subject was free to select either the 'method of principles' or the 'method of facts' (or both) in solving each problem. The information for each problem was hidden under perforated tabs, each labeled as 'Fact' or 'Principle'; and the subject was free to tear off these tabs to obtain the kind of information he believed would best help him arrive at the solution.

The BPT was administered to thirty-nine high school students in one Illinois community, for whom certain indices of educational achievement and of general aptitude were available. By means of correlational analysis, the following results were secured:

1) Those who take more mathematics courses in high school make more use of principles in the BPT ($r = .71$); they use fewer facts

($r = -.61$), even though the use of principles appears to require more time ($r = .20$).

2) While the use of facts is considered the easier method of solution, those who use facts seem less likely to obtain the correct answer ($r = -.29$).

3) Performance on the BPT, as measured by the weighted score, is a better predictor of mathematics grades than either a Language or a Non-language test of general ability.

4) Physical Science Achievement Test scores are positively correlated to a greater extent with the number of BPT problems correctly solved ($r = .61$) than is achievement on the Social Science Test ($r = .32$). However, scores in Social Science yield a lower correlation with use of BPT facts ($r = -.60$) than the scores in Physical Science ($r = -.37$).

Several distinct contributions appear to follow from our results: First, the techniques used in problem-solving emerge as psychological variables capable of serving as differential predictors of school achievement. Since individuals make characteristic use of these methods (*1*, *2*), the latter could be used in isolating the teachable elements of problem-solving. In terms of educational practice, further investigation into this area could be far reaching.

Second, the tab format of the BPT provides a situation in which the subject is able to select the kinds and amount of information he believes will best enable him to solve the problems. This technique might be extended and applied to situations where the effects of knowledge of results on subsequent test behavior can be measured as well as to obtain measures of selected personality variables (e.g., motivation, rigidity, perseveration). The effects of the ambiguity offered by the more abstract principles might possibly be extended to measure personality.

Third, the tab method might well be applied to situations where the initial selection and use of a specific series of procedures is an important aspect of the individual's subsequent training.

## REFERENCES

*1*) B. S. Bloom and Lois J. Broder. "Problem solving processes of college students." *Suppl. educ. Monogr.*, University of Chicago Press, 1950, No. 73.

*2*) E. L. Gaier. "Selected personality variables and the learning process." *Psychol. Monogr.*, 1952, 66, No. 17 (whole No. 349).

*3*) — — —. "The rôle of knowledge in problem-solving." *Progressive Education*, 1953, 30, 138–141.

# NORMATIVE TEST DATA FOR ADULT MALES OBTAINED BY HOUSE-TO-HOUSE TESTING

ELIZABETH P. HAGEN AND ROBERT L. THORNDIKE

Teachers College, Columbia University

As part of a project[1] to study the feasibility of carrying out one or more nation-wide censuses of aptitudes in the adult population, a series of pilot studies was carried out with groups of men tested at their residence units. In several of these pilot studies data were obtained for a set of three tests of ability—a vocabulary test, a mechanical principles test, and an arithmetic reasoning test. In spite of some limitations in representativeness of the samples, these norms for general adult groups appear of sufficient interest to make them available in published form.

The data were obtained during field trials in 1950 and 1951. Most of the field work was done under the direction of the project staff at Teachers College, but one set of surveys was conducted by the United States Bureau of the Census. In all of the surveys the population was defined as males eighteen years old and over. The three brief ability tests (vocabulary, mechanical comprehension, and arithmetic reasoning) were administered by the interviewer to the respondent in his own home. In the surveys conducted by the project staff, the interviewers were primarily college students—either undergraduate upper-classmen or graduate students majoring in psychology. In the surveys conducted by the Bureau of the Census, the interviewers regularly employed by the Bureau for the monthly Current Population Survey were used.

The surveys done by the project staff were carried out in Centre County, Penn.; New York, N. Y.; Springfield, Mass.; and Washington, D. C. Selection of areas was in considerable measure a matter of expediency, locations being chosen in which competent directors could be obtained for the local field work. The sampling procedure, in general, was to break the large, heterogeneous area into a number of segments relatively homogeneous in socio-economic level which were then sampled at random in proportion to the total population of the area. Blocks and dwelling units from

[1] The work reported here was done for the Human Resources Research Center, Lackland AF Base, under Contract AF 33(038)-13474.

the selected segments were drawn at random also, so that the resulting sample of units within an area was a 'probability sample'.

The survey by the Bureau of the Census was conducted in seven of the sample areas which serve as sampling units for the Bureau's Current Population Survey. These seven areas were Atlanta, Ga.; Battle Creek, Mich.; Fort Wayne, Ind.; Greenwood, S. C.; Lancaster, Penn.; New York, N. Y.; and Portland, Maine.

Although an attempt was made to locate units in different parts of the country and in different types of communities, the areas which were selected cannot be considered representative of the country as a whole. The most notable omission is the lack of communities west of the Mississippi River and in Southern rural areas. However, the data obtained in the surveys are believed to be of some interest in showing (a) the general level of performance on the tests, (b) the order of magnitude of community differences; and (c) the relationships between education, socio-economic status, and age and performance on the tests.

## MATERIALS

*Multiple Choice Vocabulary Test.*—The test was a twenty-word test made up of two words from the vocabulary subtest of each of ten levels of the IER Intelligence Scale CAVD.[2] It is a steeply graded power test in which almost everyone can get a score, but only a very few get perfect scores. The test was presented to the respondent on a card. The respondent had to read each word and the five choices for himself and give the number of the correct answer. The test was continued until the respondent failed five consecutive items or until he refused to continue. The score was the number of correct responses.

*Mechanical Comprehension Test.*—This test consisted of the following eighteen items chosen from Form A of the Differential Aptitude Tests, Mechanical Comprehension Test: 1, 4, 9, 12, 17, 20, 25, 28, 33, 36, 41, 44, 49, 52, 57, 60, 65, 68. Items were cut out and mounted on 5 x 8 cards. They were shown to the respondent one at a time, and, except in the case of refusals, each respondent took all eighteen items. In cases where the respondent could not read, the items were read to him. The test was scored number right minus one half the number wrong.[3]

[2] Robert L. Thorndike, "Two screening tests of verbal intelligence," *J. Appl. Psychol.*, 1942, 26, 128–135.

TABLE 1.—TEST RELIABILITIES AND INTERCORRELATIONS

| Test | Mechanical | Arithmetic | Reliability |
|---|---|---|---|
| Vocabulary | .32 | .59 | .85 |
| Mechanical Comp. | | .47 | .59 |
| Arithmetic Reas. | | | .83 |

*Arithmetic Reasoning Test.*—This was a twelve-item test steeply graded in difficulty. The test is reproduced in Appendix A. Ten of the problems were chosen from materials gathered in the course of preparing the IER Intelligence Scale CAVD. Two easier problems were added in an attempt to reduce resistance to the test and the piling up of zero scores. Items were presented one at a time on 5 x 8 cards; subjects could use scratch paper; and the answers obtained were reported to the interviewer. As in the mechanical test, items were read to respondents who were unable to read. The score was the number of items answered correctly.

Table 1 shows split-half reliabilities and intercorrelations of the tests for 1060 cases included in one survey by project personnel.

## RESULTS

Tables 2, 3, and 4 give the percentile ranks of the scores on each test and the means and standard deviations on each test for a composite of all of the communities in the surveys done by the Bureau of the Census and a similar composite for the communities surveyed by Teachers College. These tables also show the percentile ranks of the scores for the Teachers College surveys when the sample is weighted and adjusted for age and education according to the 1950 U. S. Bureau of the Census reports for the country as a whole.

On the arithmetic and mechanical comprehension tests, the percentile ranks of the raw scores are closely similar for both the Bureau of the Census and the Teachers College surveys. On the vocabulary test, the percentile ranks are quite similar for the lower half of the distribution but show somewhat less agreement for the upper half of the distribution.

However, the percentile ranks of the weighted sample are quite

[3] The courtesy of the Psychological Corporation in permitting the use of this material is gratefully acknowledged.

TABLE 2.—PERCENTILE EQUIVALENTS OF RAW SCORES ON VOCABULARY TEST FOR ALL COMMUNITIES

| Raw Score | Percentile Rank | | |
|---|---|---|---|
| | Bureau of Census Survey | Teachers College Survey | |
| | | Unweighted | Weighted* |
| 0 | 1 | 1 | 2 |
| 1 | 2 | 2 | 6 |
| 2 | 3 | 4 | 8 |
| 3 | 4 | 5 | 12 |
| 4 | 6 | 8 | 16 |
| 5 | 9 | 12 | 23 |
| 6 | 14 | 19 | 31 |
| 7 | 22 | 28 | 41 |
| 8 | 32 | 37 | 51 |
| 9 | 45 | 45 | 60 |
| 10 | 58 | 53 | 67 |
| 11 | 69 | 61 | 74 |
| 12 | 78 | 69 | 80 |
| 13 | 85 | 76 | 85 |
| 14 | 91 | 82 | 89 |
| 15 | 95 | 87 | 93 |
| 16 | 97 | 92 | 96 |
| 17 | 98 | 95 | 98 |
| 18 | 99.2 | 98 | 99 |
| 19 | 99.6 | 99 | 99.7 |
| 20 | 99.9 | 99.9 | |
| N | 597 | 2292 | |
| $\bar{X}$ | 9.3 | 9.7 | 8.2 |
| SD | 3.4 | 4.2 | 4.3 |

* Weighted for age and education.

different from the other two. In all three tests in both the Census surveys and the Teachers College surveys, there appeared to be an under-representation of the lower educational levels and over-representation of the higher educational levels. This was probably due to (a) the biases introduced by refusals to take the test in the over-forty-five age group and (b) inclusion of communities in the study which had a higher level of education than the country as a whole.

Table 5 shows the 10th, 50th, and 90th percentiles for each of the communities on each of the tests. As can be seen from the table,

TABLE 3.—PERCENTILE EQUIVALENTS OF RAW SCORES ON ARITHMETIC TEST FOR ALL COMMUNITIES

| Raw Score | Percentile Rank | | |
|---|---|---|---|
| | Bureau of Census Survey | Teachers College Survey | |
| | | Unweighted | Weighted* |
| 0 | 1 | 1 | 2 |
| 1 | 3 | 3 | 6 |
| 2 | 7 | 7 | 12 |
| 3 | 15 | 14 | 23 |
| 4 | 27 | 27 | 39 |
| 5 | 40 | 40 | 53 |
| 6 | 52 | 50 | 63 |
| 7 | 65 | 61 | 73 |
| 8 | 78 | 72 | 82 |
| 9 | 87 | 82 | 89 |
| 10 | 94 | 90 | 94 |
| 11 | 98 | 96 | 98 |
| 12 | 99.7 | 99 | 99 |
| N | 600 | 2276 | |
| $\bar{X}$ | 5.8 | 5.9 | 5.1 |
| SD | 2.6 | 2.8 | 2.8 |

* Weighted for age and education.

certain communities, notably Springfield, Mass., and Lancaster, Penn., tend to rank high on all of the tests. However, most of the communities tend to perform better on one test than on the other two. Rural areas, such as Centre County, Penn., and Greenwood, S. C., tend to be lower than other communities on the verbal test, but do relatively much better on the mechanical test.

Table 6 shows the results for each test according to socio-economic status. Socio-economic status was determined from the weighted sum of three variables. The three variables used were occupation, house type, and neighborhood rating. These were combined according to procedures described by Warner.[4] An 'A' rating represents highest socio-economic status; a 'D', lowest.

The increase in median score from the lowest socio-economic

[4] W. L. Warner, M. Meeker, and K. Eells, *Social Class in America*. Chicago: Science Research Associates, 1949.

TABLE 4.—PERCENTILE EQUIVALENTS OF RAW SCORES ON MECHANICAL COMPREHENSION TEST FOR ALL COMMUNITIES

| Raw Score | Percentile Rank | | |
|---|---|---|---|
| | Bureau of Census Survey | Teachers College Survey | |
| | | Unweighted | Weighted* |
| −1½ and below | 1 | 1 | 1 |
| 0 | 2 | 2 | 4 |
| 1½ | 4 | 4 | 7 |
| 3 | 8 | 8 | 12 |
| 4½ | 13 | 14 | 19 |
| 6 | 22 | 24 | 30 |
| 7½ | 36 | 36 | 44 |
| 9 | 52 | 50 | 58 |
| 10½ | 66 | 64 | 71 |
| 12 | 78 | 76 | 82 |
| 13½ | 88 | 88 | 91 |
| 15 | 94 | 92 | 96 |
| 16½ | 97 | 97 | 98 |
| 18 | 99 | 99 | 99 |
| N | 610 | 2278 | |
| $\bar{X}$ | 8.8 | 8.9 | 8.1 |
| SD | 3.9 | 4.1 | 4.2 |

* Weighted for age and education.

group to the highest was about 1.8 standard deviation units (based on the standard deviation of the total group) in the case of the vocabulary test, 1.4 standard deviation units for the arithmetic test, and 1.0 standard deviation units for the mechanical comprehension test. In each case, there was a regular progression in score from level to level, each rise in socio-economic level being accompanied by a rise in score. It is perhaps worth noting that for the vocabulary test the level A-level B difference is the largest one, while for the other two tests it is the smallest one. That is, the highest socio-economic group is particularly characterized by a high verbal level. These findings for adult groups are consistent with results from many studies of the relationship between father's occupation or family socio-economic status and the scores made by children on intelligence tests.

Table 7 shows the distribution of scores by educational level for

TABLE 5.—SELECTED PERCENTILE SCORES ON THE VOCABULARY TEST, MECHANICAL COMPREHENSION TEST, AND ARITHMETIC TEST FOR COMMUNITIES IN THE TEACHERS COLLEGE AND BUREAU OF THE CENSUS SURVEYS

| | Vocabulary | | | Mech. Comp. | | | Arithmetic | | |
|---|---|---|---|---|---|---|---|---|---|
| | $P_{10}$ | $P_{50}$ | $P_{90}$ | $P_{10}$ | $P_{50}$ | $P_{90}$ | $P_{10}$ | $P_{50}$ | $P_{90}$ |
| *Teachers College Surveys* | | | | | | | | | |
| Springfield, Mass. | 6.8 | 11.7 | 16.2 | 4.7 | 10.4 | 15.9 | 3.5 | 7.2 | 10.4 |
| New York, N. Y. | 5.9 | 11.4 | 16.4 | 3.1 | 8.4 | 13.7 | 3.0 | 6.6 | 10.0 |
| Washington, D. C. | 3.7 | 9.2 | 15.0 | 2.4 | 8.6 | 14.4 | 1.9 | 5.4 | 10.2 |
| Centre Co., Pa. | 3.9 | 7.4 | 12.3 | 4.5 | 9.6 | 14.2 | 2.4 | 5.4 | 9.1 |
| *Bureau of the Census Surveys* | | | | | | | | | |
| Atlanta, Ga. | 3.4 | 8.9 | 13.4 | 2.9 | 8.2 | 14.3 | 2.7 | 5.8 | 9.1 |
| Battle Creek, Mich. | 5.1 | 8.7 | 12.6 | 4.1 | 9.3 | 14.4 | 2.6 | 5.2 | 8.9 |
| Fort Wayne, Ind. | 6.0 | 9.7 | 13.7 | 5.3 | 9.2 | 13.7 | 2.7 | 6.3 | 9.4 |
| Greenwood, S. C. | 2.9 | 8.2 | 12.9 | 4.1 | 8.6 | 13.5 | 1.7 | 4.2 | 8.4 |
| Lancaster, Pa. | 6.8 | 9.6 | 14.1 | 5.4 | 9.1 | 13.8 | 3.9 | 6.8 | 9.7 |
| New York, N. Y. | 5.0 | 11.5 | 15.0 | 2.4 | 7.7 | 13.6 | 2.7 | 5.9 | 8.4 |
| Portland, Maine | 5.2 | 9.3 | 12.2 | 3.1 | 8.2 | 13.7 | 2.1 | 6.1 | 10.0 |

TABLE 6.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS, BY SOCIO-ECONOMIC STATUS

| Percentile and Test | Socio-Economic Statue | | | |
|---|---|---|---|---|
| | D | C | B | A |
| Vocabulary | | | | |
| 10th percentile | 1.9 | 5.2 | 6.7 | 9.1 |
| 25th percentile | 4.5 | 6.7 | 9.1 | 12.0 |
| 50th percentile | 6.8 | 9.2 | 11.9 | 14.5 |
| 75th percentile | 9.2 | 11.9 | 14.5 | 16.4 |
| 90th percentile | 11.4 | 14.4 | 16.4 | 17.9 |
| Mechanical Comprehension | | | | |
| 10th percentile | 1.8 | 3.5 | 5.1 | 6.1 |
| 25th percentile | 4.7 | 7.7 | 7.6 | 8.6 |
| 50th percentile | 7.3 | 8.9 | 10.3 | 11.4 |
| 75th percentile | 10.1 | 11.5 | 12.8 | 14.2 |
| 90th percentile | 12.4 | 15.3 | 15.5 | 16.3 |
| Arithmetic | | | | |
| 10th percentile | 1.3 | 2.7 | 4.1 | 5.1 |
| 25th percentile | 2.6 | 3.9 | 5.7 | 6.8 |
| 50th percentile | 3.9 | 5.7 | 7.6 | 8.8 |
| 75th percentile | 5.8 | 7.8 | 9.3 | 10.4 |
| 90th percentile | 7.6 | 9.3 | 10.6 | 11.3 |

TABLE 7.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS, BY YEARS OF EDUCATION COMPLETED

| Percentile & Test | Last School Grade Completed | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1–4 | 5–6 | 7 | 8 | 9 | 10–11 | 12 | 13–15 | 16 | Over 16 |
| Vocabulary | | | | | | | | | | |
| 10th percentile | 0.3 | 0.8 | 0.9 | 3.1 | 4.0 | 4.9 | 6.0 | 7.4 | 9.2 | 10.2 |
| 25th percentile | 0.9 | 2.6 | 2.8 | 5.0 | 5.8 | 6.3 | 7.5 | 9.4 | 11.8 | 12.9 |
| 50th percentile | 3.3 | 5.1 | 5.5 | 6.4 | 7.4 | 8.3 | 9.7 | 11.5 | 13.9 | 15.1 |
| 75th percentile | 4.7 | 7.0 | 7.9 | 8.8 | 10.0 | 11.1 | 12.0 | 13.9 | 15.7 | 16.9 |
| 90th percentile | 7.8 | 9.2 | 9.4 | 11.2 | 12.3 | 13.0 | 14.4 | 15.8 | 17.0 | 18.2 |
| Mechanical Comp. | | | | | | | | | | |
| 10th percentile | −1.8 | −0.95 | 1.4 | 2.5 | 2.4 | 2.7 | 4.3 | 5.0 | 5.1 | 7.0 |
| 25th percentile | 2.9 | 3.4 | 4.2 | 4.4 | 5.7 | 5.4 | 6.9 | 7.0 | 7.9 | 8.9 |
| 50th percentile | 6.6 | 6.0 | 6.5 | 7.6 | 8.2 | 8.0 | 9.4 | 10.1 | 10.2 | 11.6 |
| 75th percentile | 9.4 | 8.9 | 9.2 | 11.8 | 10.7 | 10.7 | 11.8 | 12.5 | 14.5 | 13.9 |
| 90th percentile | 11.1 | 12.4 | 11.8 | 12.3 | 12.6 | 13.0 | 19.2 | 15.1 | 16.4 | 16.3 |
| Arithmetic | | | | | | | | | | |
| 10th percentile | 0.5 | 0.0 | 0.8 | 1.6 | 2.1 | 2.6 | 3.3 | 4.0 | 5.0 | 5.8 |
| 25th percentile | 1.3 | 1.6 | 2.1 | 2.8 | 3.2 | 3.7 | 4.5 | 5.3 | 7.2 | 7.3 |
| 50th percentile | 2.8 | 2.8 | 3.2 | 3.9 | 4.4 | 4.9 | 6.3 | 7.4 | 8.8 | 9.0 |
| 75th percentile | 3.8 | 4.2 | 4.6 | 5.5 | 6.6 | 7.0 | 8.3 | 9.1 | 10.3 | 10.4 |
| 90th percentile | 4.7 | 5.0 | 6.0 | 7.1 | 8.1 | 8.4 | 9.4 | 10.3 | 11.2 | 11.4 |

each of the three tests. The vocabulary and arithmetic tests both show steep and steady increases in median score as the level of education increases. The mechanical comprehension test shows an increase in median score as educational level increases, but the increase is neither as steady nor as large as that for the other two tests. On the vocabulary test, the increase in median score from the lowest to the highest educational level is equal to 2.8 standard deviations; on the arithmetic reasoning test it is equal to 2.2 standard deviations, and on the mechanical test to 1.2 standard deviations. These relationships between education and scores on the tests—especially the vocabulary and arithmetic tests—are consistent with the correlation of .70 which was reported by Tuddenham for the relationship between education and AGCT scores.[5]

Table 8 shows the distribution of scores for the three tests for

[5] Read D. Tuddenham, "Soldier intelligence in World Wars I and II," *The American Psychologist*, 1948, 3, 54–56.

TABLE 8.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS BY WEIGHTED AGE GROUPS*

| Percentiles and Tests | Age Groups | | | |
|---|---|---|---|---|
| | 18–25 | 26–35 | 36–45 | Over 45 |
| Vocabulary | | | | |
| 10th percentile | 2.7 | 2.4 | 2.5 | 1.3 |
| 25th percentile | 4.7 | 5.2 | 5.1 | 4.3 |
| 50th percentile | 6.7 | 7.9 | 7.8 | 7.3 |
| 75th percentile | 9.4 | 10.8 | 11.4 | 11.0 |
| 90th percentile | 11.5 | 13.6 | 14.0 | 14.3 |
| Mechanical Comprehension | | | | |
| 10th percentile | 0.8 | 2.2 | 2.3 | 2.7 |
| 25th percentile | 5.0 | 5.2 | 5.4 | 5.3 |
| 50th percentile | 7.9 | 8.3 | 8.2 | 7.6 |
| 75th percentile | 10.7 | 10.8 | 10.8 | 10.9 |
| 90th percentile | 13.5 | 13.2 | 13.4 | 13.2 |
| Arithmetic | | | | |
| 10th percentile | 0.5 | 1.4 | 1.3 | 1.3 |
| 25th percentile | 2.4 | 2.8 | 2.7 | 2.6 |
| 50th percentile | 4.0 | 4.4 | 4.3 | 4.1 |
| 75th percentile | 6.5 | 6.8 | 6.9 | 6.5 |
| 90th percentile | 8.5 | 8.6 | 8.7 | 8.5 |

Age groups weighted for educational level according to 1950 Bureau of the Census data on education for the country as a whole.

age groups which have been weighted according to 1950 Bureau of the Census data on educational level for different age groups for the country as a whole. The weighting was made to correct for educational biases in the sample. In the sample, the higher educational levels in the eighteen to twenty-five year old group were under-represented, presumably because a large number of these men were away from home at the time of the survey and were not available for interviews. In the older age groups, the lower educational levels are under-represented, in part, at least, because the older men with less education tended to refuse to take the tests.

Before the age groups were weighted for educational level, the increase in test scores by age followed the pattern for education. After the correction, the median scores on all three tests rise slightly in the twenty-six to thirty-five year old group, then drop off in the thirty-six to forty-five and over-forty-five age groups. In the vocab-

ulary and arithmetic tests, the median score for the upper two age groups is slightly above that for the eighteen to twenty-five age group. In the mechanical comprehension test, the median score for the over-forty-five age group is slightly below that for the eighteen to twenty-five year old group. However, all of the differences in median scores are small—the largest being approximately .3 of a standard deviation unit.

## APPENDIX A

### *Arithmetic Reasoning Test*

1) If 2 pencils cost 5¢, what will 4 pencils cost?

2) You buy a newspaper for 5¢ and a pack of cigarettes for 21¢. If you give the clerk 50¢, how much change should you get?

3) What does a pound of candy cost when you pay 10¢ for ¼ of a pound?

4) A man bought land for $400. He sold it for $445, gaining $15 an acre. How many acres were there?

5) How much is ⅞ of 28?

6) 20 = how many times 12?

7) How many times as long as 8 feet is 12 yards?

8) Counting that 25 dozen sheets of paper are worth 10¢, how many sheets of paper are worth ⅕ of a cent?

9) How many quarters of a quarter is half of a half?

10) How long will enough food for 400 men for 120 days last 250 men?

11) A man started with $12,500 and doubled his capital every year for five years. How much had he at the end of the fifth year?

12) A salesman was to receive 10¢ for every dollar's worth of merchandise he sold, an extra 10¢ for every dollar over 100, and a third 10¢ for every dollar over 1000. The salesman sold $P$ plus $1000 worth of merchandise. How much should he receive?

# PREDICTING ATTRITION-SURVIVAL IN FIRST YEAR ENGINEERING

JOHN P. MALLOY, BOLESLAW WYSOCKI, AND LEO F. GRAHAM

Marquette University

During recent years Marquette University engineering students have been administered the Minnesota Paper Form Board Form AA, as part of a battery of entrance tests. The American Council on Education, 1949 edition, the Coöperative Mathematics Test Form Y, and the Coöperative Reading Test Form RX, also have been included in the battery. There has been some question about the effectiveness of the Minnesota Paper Form Board. Some advisors have felt that the intercorrelations between the Minnesota Paper Form Board and other tests in the battery, particularly the ACE-Q score, were such that relatively little was gained by its inclusion. This study was undertaken in order to evaluate the importance of the Minnesota Paper Form Board as a predictor of success in engineering college.

A second purpose of this study was to determine the effectiveness of each test for predicting survival in Marquette's Engineering College. The majority of the studies investigating the prediction of college success by means of psychological measures have utilized regular correlation methods with college grades as the criterion. Since survival in the engineering curriculum ordinarily implies entrance in the profession, it was felt that survival was an appropriate and logical criterion. Although survival for the full four years of work would be a better criterion, survival for the first year was used because of the unavailability of longitudinal data encompassing a longer period of time.

Minnesota Paper Form Board scores, A.C.E. scores, and Coöperative Mathematics scores were available for one hundred and ninety-four students who had enrolled in Marquette Engineering College in the fall of 1952. The sample represented seventy-one per cent of the class. Although the reading test scores were available, they were not utilized as part of this study. The MPB scores were unavailable for the remaining twenty-nine per cent of the class and hence this group could not be included in the study.

Of the one hundred and ninety-four students, one hundred and

sixty-one survived the first year. Causes for drop-out were disregarded and no distinction was made betweeen students that were dropped by the college and those who dropped of their own accord for various reasons.

The attrition rate in this sample was eighteen per cent. An examination of the entire class revealed that thirty-one per cent of the students failed to remain the entire year. The difference between the sample's attrition rate and that for the entire class probably can be traced in part to ramifications of Marquette's admission policy.

The university regularly accepts high school graduates who rank in the top two-thirds of their class without further qualifications. If a student cannot meet this standard, it is possible for him to gain entrance provided that he performs satisfactorily on the freshman test battery. This battery includes the tests previously mentioned except for the Minnesota Paper Form Board. It also includes the Coöperative Mechanics of Expression Test. All students whose scores were used in this study were those admitted on the basis of satisfactory high school performance. For this reason they were a select group.

From the differences of the rate of attrition between those students admitted on the basis of high school rank and those who failed to meet the standard but are later admitted on the basis of their performance on the psychological tests, it would seem that the latter group have a much poorer chance for success than those that are admitted without qualification. At the present time the admission policies of the University are undergoing scrutiny as part of a self-survey now being conducted throughout the University by the administration and by the faculty.

## TREATMENT OF THE DATA

The usual method for determining the predictive effectiveness of psychological measures has been the utilization of correlation methods where both independent and dependent variables are numerical variables. Since the attrition-survival criterion results in a dichotomy, the usual regression techniques were inappropriate for this study and another design was necessary.

Discriminant analysis and the serial correlation which can result from it satisfy this need. Fisher (*1*) developed the former technique which basically involves ascertaining appropriate weights for a series of variables yielding maximum separation in two groups, each

TABLE I.—MEAN SCORES FOR THE SURVIVAL AND ATTRITION GROUPS

| Numerical Variable | Survival N-161 | Attrition N-33 | Difference |
|---|---|---|---|
| ACE Q ($x_1$) | 53.31 | 47.12 | 6.19 |
| ACE L ($x_2$) | 67.27 | 61.45 | 5.82 |
| MPB ($x_3$) | 48.36 | 42.42 | 5.84 |
| Math ($x_4$) | 25.48 | 21.88 | 3.60 |

of which is assumed to be normally distributed. Wert, Neidt and Ahmann (*2*) have illustrated the use of this design for psychological data. Accordingly the data in this study were analyzed with discriminant analysis except where the computation of Pearson Product Moment r's was appropriate.

## RESULTS

An inspection of Table I provides a review of the performance of the attrition-survival groups. The survival groups performance surpasses the attrition group in each instance. The solution of the coefficients provided the function:

$$V = .00026832x_1 + .00006118x_2 + .00027178x_3 + .00003033x_4$$

where

$V$ = deviation relative score

$x_1$ = deviation ACE-Q score $x_3$ = deviation MPB score

$x_2$ = deviation ACE-L score $x_4$ = deviation Math score

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN VARIABLES

N = 194

| Survival $R_{bis}$ | $Q_{(x_1)}$ | $L_{(x_2)}$ | $MPB_{(x_3)}$ | $Math_{(x_4)}$ |
|---|---|---|---|---|
| | .398 | .239 | .375 | .258 |
| Q (r) | | .408 | .386 | .444 |
| L (r) | | | .163 | .371 |
| MPB (r) | | | | .341 |

$R_{bis(x_1x_2x_3x_4)} = .473$
$R_{bis(x_1x_2x_3)} = .472$
$R_{bis(x_2x_3x_4)} = .424$
$R_{bis(x_1x_2x_4)} = .413$
$R_{bis(x_1x_3x_4)} = .446$

$R_{bis(x_1x_2)} = .407$
$R_{bis(x_1x_3)} = .465$
$R_{bis(x_1x_4)} = .408$
$R_{bis(x_2x_3)} = .416$
$R_{bis(x_2x_4)} = .301$

$R_{bis(x_3x_4)} = .399$

TABLE III.—TESTS OF SIGNIFICANCE RESULTING WHEN ONE TEST WAS DELETED FROM A COMBINATION OF TESTS PREDICTING THE CRITERION

| Predictor Combination ($R_{bis}$) | Test Eliminated | F | d.f |
|---|---|---|---|
| $X_1X_2X_3X_4$ | $X_4$ | .1 | 1,189 |
| $X_1X_2X_3X_4$ | $X_3$ | †5.6* | 1,189 |
| $X_1X_2X_3X_4$ | $X_2$ | 2.4 | 1,189 |
| $X_1X_2X_3X_4$ | $X_1$ | 4.2* | 1,189 |
| $X_1X_2X_3$ | $X_3$ | 5.5* | 1,190 |
| $X_1X_2X_3$ | $X_2$ | .7 | 1,190 |
| $X_1X_3$ | $X_3$ | 5.6* | 1,191 |

* Significant at the five per cent level.

† This is interpreted to mean that a significant loss in prediction effectiveness results at five per cent confidence level when this test is eliminated from the battery.

From a review of Table II the predictive effectiveness of each variable may be observed. The zero order Pearson r's between each pair of numerical variables were also computed. A test of the significance of the difference from zero of the smallest biserial r, that between L and survival, was made. It was highly significant. The smallest Pearson zero order r, that between the ACE-L score and the MPB, was significant at the five per cent level but not at the one per cent level.

To determine the best predictive combination the discriminant functions were computed in such a fashion as successively to delete each of the prediction variables. The resulting F-tests are recorded in Table III. The L-score and the Coöperative Mathematics Test may be eliminated from the prediction equation since neither contributes significantly. It is obvious that the Q score and the MPB score make the best two variable predictors.

## DISCUSSION

A surprising outcome of this study was the ineffectiveness of the mathematics score in the prediction of attrition-survival. It is possible that an extension of this study to include students dropping out in the sophomore year might reveal an increase in the importance of this variable. It is well known that the calculus is a major stumbling block for embryonic engineers.

The importance of the Minnesota Paper Form Board may result from the inclusion of engineering drawing and descriptive geometry

in the freshman curriculum. It seems certain that it should be retained in the test battery for engineering students.

As an outcome of this study a table with a two-way entry was constructed, using the ACE-Q score and the MPB score. This table was fashioned so that it can be entered with the percentile scores of both the ACE-Q and MPB scores, and the chances in one hundred of a student surviving is determined. When one enters this table with the Q and MPB scores at the fifth percentile, a student's chance of surviving is sixty-one in 100. If a student achieves scores at the 95th percentile on both measures, his chances for survival for the first year of engineering at Marquette University are ninety in one hundred.

## CONCLUSIONS

1) The ACE-Q score and the Minnesota Paper Form Board score were the best predictors of survival in first year engineering at Marquette University.

2) These two measures are relatively effective as predictors of attrition-survival.

3) The Coöperative Mathematics Test contributed least to prediction of the criterion.

4) Since a legitimate question can be raised as regards the representativeness of the sample when compared with the entire class, caution must be used in generalizing from this study.

## SUMMARY

A study of attrition-survival was made at the Marquette College of Engineering. By means of discriminant analysis and serial correlation it was discovered that the ACE-Q score and the Minnesota Paper Form Board were comparatively effective as predictors of survival. A multiple biserial r equal to .473 resulted when four predictor variables were used. When the two best cited tests were used, a serial correlation coefficient of .465 was obtained.

## REFERENCES

1) R. A. Fisher. "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7: 179–188, 1936.

2) J. E. Wert, C. O. Neidt, and J. S. Ahmann. *Statistical Methods in Educational and Psychological Research*. New York: Appleton-Century-Crofts, Inc., 1953.

# BILINGUALISM AND BICULTURALISM

JAMES P. SOFFIETTI

Syracuse University

The literature on bilingualism (*1, 2, 4, 5, 6*) is voluminous. It deals chiefly with its effects on the mental and linguistic development of the child and his school achievement, as well as his personal and social adjustment. As a corollary, bilingualism also raises the question as to the advisability of teaching a second language in our schools and that of the time and method of such instruction.

These problems have been discussed and investigated from many points of view both in this country and abroad. On one sole point there seems to be general agreement: that the findings so far are inconclusive and that the problem needs further careful investigation. The general implication seems to be that bilingualism is a much more complicated condition of affairs than that of the use of two languages by an individual.

If bilingualism is not a simple concept, then it needs further differentiation and clarification. From the point of view of dictionary definitions, the concept of bilingualism is quite simple. The American College Dictionary defines it (1) as 'the habitual use of two languages'; and (2) 'the ability of being bilingual'. It defines the bilingual as a person 'able to speak one's native language and another with approximately equal facility'.

The complications begin to arise when one attempts to apply the dictionary definition to specific cases. It then loses most of its value as a working concept. One soon discovers that most of the practical situations referred to as 'bilingual' involve factors that extend far beyond those of the 'habitual use of two languages'. Investigators have usually been aware of this, but somehow have not realized the necessity of isolating such factors from the basic concept of bilingualism.

Thus we find the researcher talking about 'bilingualism' when he is referring not only to the use of two languages, but also to the fact (and at times mainly so) that the individual or group under investigation is living in a different 'cultural atmosphere' at home or in his neighborhood from the one prevailing in the community at large. In other words, to the investigator bilingualism refers not

only to two distinct patterns of 'linguistic habits', but also to distinct patterns of 'cultural habits' in all of their anthropological meaning. His problem is not merely one of bilingualism, but also one of what we may justifiably call 'Biculturalism'.

Many students of bilingualism appear to assume that the concept of language already comprises the factors of cultural differences. This is a misconception. While it is true that language behavior is an integral part of cultural behavior and that the latter would be impossible without its linguistic components, it does not mean that to a given language system there has to correspond a specific cultural structure or area. It is possible for otherwise totally unrelated cultures to share in one linguistic system and for totally unrelated languages to share in one cultural system. Language, race, and culture are not necessarily correlated (*9*). Even the fact that language tends to canalize the speaker's response does not mean that it is inseparably bound to a specific culture (*8*).

By keeping a clearcut distinction between the concepts of 'bilingualism' and 'biculturalism', the investigator will find it easier to outline his problem of research, conduct his investigation, categorize his findings and report the facts; and the classroom teacher will find it easier to solve his 'bilingual' problems. But before discussing these two concepts in greater detail, there is another important one that should be isolated from the current meaning of the term bilingualism.

Researchers have also been aware of the fact that there are degrees of bilingualism. For various reasons, not all bilinguals are proficient to the same extent in the two languages used. In such cases we usually speak of 'linguistic accents'. They are due to various forms of interference between the patterned habits of the mother tongue and those of the second language with results of distortion, substitution, or omission, of speech sounds or grammatical structures, or both.

To the same extent we may speak of 'Cultural Accents'. This term is not intended to be a metaphorical expression. 'Cultural accents' are real and of the same nature as the linguistic ones. They, too, are the result of conflicting habit patterns persistently interfering with each other and resulting in any number of distortions or outright substitutions. They may be detected at any point and in various degrees throughout the daily routine activities of a non-native member of a society: from table manners to non-

UNIVERSITY OF MICHIGAN

verbal forms of greeting; from gait to ways of driving a car; from choice of foods to manner of preparing them; from little things about dressing to features in house furnishing; from attitudes towards members of the family to attitudes towards one's 'superiors' in business, and in intellectual or socio-economic status. They can be perceived in non-verbal expressions of emotion as well as in the value system of the individual under consideration.

At times, however, a cultural accent may be so subtle that all we can say is that there is something 'foreign' about the person in question without our being able to put our finger on the distinguishing feature involved.

Learning to behave in conformity with the non-verbal cultural patterns of another society can be as much of a task as learning its verbal patterns, i.e., its language. To get rid of cultural accents can be just as difficult as getting rid of linguistic ones: both are based on deep-rooted systems of habit, some of which have strong emotional basis.

The concepts of linguistic and cultural accents can play an important rôle in the definition of certain types of minority groups. In many instances the only distinguishing feature of a minority is the presence of a certain degree or type of cultural or linguistic accents, or both, which have survived as vestiges through any number of generations, while conformity exists in all the other aspects of their societal behavior.

By keeping a clearcut distinction among the concepts that have been developed above, instead of gathering them under the single label of bilingualism, the research worker will be able to distinguish between the following four basic types of situations:

(1) bicultural-bilingual
(2) bicultural-monolingual
(3) monocultural-bilingual
(4) monocultural-monolingual

We are omitting from this discussion the possibility of multilingual and multicultural situations, as well as further differentiations within the bicultural and bilingual concepts as suggested, e.g., by Mary R. Haas and others. (*7*).

(1) The bicultural-bilingual person. He is the one that is usually considered as the true 'bilingual'. He is able, or actually does, participate intimately in two cultures (*3*). An example would be the child of immigrant parents who at home continue to speak the

native tongue and maintain most of the native customs, beliefs and value patterns.

(2) A bicultural-monolingual case would be that of a child of immigrant parents who, for a number of reasons, have given up speaking their native tongue or whose native language happens to be English (such as a British or Irish family), but who have not given up their native customs, beliefs or value patterns.

(3) A monocultural-bilingual example is that of a child who participates in one sole culture, but who has learned to use a second language either in school or at home. This would be the case, e.g., when one or both of the parents are bilingual and have decided that at least one of them shall always speak to the child in the foreign language, although the remaining aspects of cultural behavior are those of the community at large.

(4) It goes without saying, that the most common situation in the United States is the monocultural-monolingual one. In this case, the child does not have to contend with linguistic and cultural conflicts. (We are omitting from this paper the discussion of problems of sub-cultural nature.)

In the first three types of situations, the experienced investigator would detect more often than not various degrees of foreign accent in either the language or cultural patterns, or both. Perhaps cases of perfect bilingualism and biculturalism are extremely rare, if at all possible.

If one studies the literature dealing with the effects of bilingualism on the intellectual and social development of the child in the light of the above new concepts, he will find that most of the difficulties and retardations indiscriminately ascribed to bilingualism are rather due to the bicultural aspects of the situation under consideration. It is the living in two distinct cultures either overtly or in one's internal life that might create adjustmental problems. It is a conflict between ways of life, beliefs, customs, value systems and not necessarily one between language systems.

A person learning a second language in a monocultural setting will not automatically learn a whole new set of cultural patterns and develop cultural conflicts or accents. As a matter of fact, it is seldom that a person studying a second language in our schools ever gets beyond the 'synonym stage' (i.e., beyond the purely bilingual stage as defined in this paper) of foreign language learning: all he usually learns is another way of referring to the same or simi-

UNIVERSITY OF MICHIGAN LIBRARIES

lar things, situations, or events. 'Mère' is just another word for 'mother'; 'je vais chez moi' another way of saying 'I am going home'. Rarely do his meanings extend beyond the simple denotation to include emotional and cognitive associations common to all the participants in the culture where that particular language is spoken. It is only by living in the culture where the language is spoken that the full meanings of its utterances are acquired. Learning French or German, or Urdu, in a monocultural setting will no more create problems of intellectual and social nature than learning Esperanto.[1]

In conclusion, it is hoped that the concepts of 'biculturalism' and 'cultural accents' developed in this article will aid the educator, the psychologist and the sociologist in their effort to isolate the distinctive elements that characterize the so-called bilingual person or situation. It is also hoped that these concepts will help the bicultural person to a better understanding of himself.

REFERENCES

1) Seth Arsenian. "Bilingualism in the postwar world." *The Psychological Bulletin*, 1945 (February), pp. 65–86.

2) — —. *Bilingualism and Mental Development*. New York: Teachers College, Columbia University, 1937.

3) J. S. H. Bossard. "The bilingual as a person—linguistic identification with status." *American Sociological Review*, 1945, 10: 699–709.

4) Einar Haugen. *A Bibliography of Bilingualism*. Madison: University of Wisconsin. (in preparation.)

5) T. Manuel Herschel. "Bilingualism." *Encyclopedia of Educational Research*, 1950 (rev. ed.)

6) Werner F. Leopold. *Bibliography of Child Language*. Evanston: Northwestern University Press, 1952, p. 115.

---

[1] The above is in no way meant to minimize the cultural value of studying a foreign language. It will generally be agreed, however, that these values can only be gotten when the foreign language class is enriched by illustrated and dramatized discussions about the foreign culture and by comparisons of important aspects of it and its language with corresponding features in ours. As every anthropologist well knows, without some knowledge of the language it is impossible to gain a thorough understanding of a culture for the simple reason that many of the cultural traits depend on the language components for their very existence and configuration. How could anyone on his own even begin to understand our way of life without having some knowledge of English?

7) Claude Levi-Strauss, et al. "Results of the conference of anthropologists and linguists." *Memoir No. 8, I.J.A.L.*, 1953 (April), p. 42.

8) Eric H. Lenneberg. "Cognition and Ethnolinguistics." *Language*, 1953, 29: 463–71.

9) Edward Sapir. *Language*. New York: Harcourt, Brace and Co., 1921, p. 221 ff.

UNIVERSITY OF MICHIGAN LIBRARIES

# WAR-TIME ACCELERATES TEN YEARS AFTER

MARIE A. FLESHER and S. L. PRESSEY

The Ohio State University

During World War II acceleration was the most hotly debated of all educational topics, and most colleges and technical schools had some form of accelerated program. When the war ended, most of these programs seemed quickly to have been dropped, and most educators seemed desirous to forget the topic as quickly as possible. However, the Korean war and likely indefinite continuance of the draft or related measures in view of continuing international tensions, makes the matter again of at least potential importance. Moreover, the few studies which have been made of war-time programs of acceleration have shown not the anticipated unfortunate effects, but rather the reverse! Thus the writers studied intensively the one hundred and four women who in the four colleges of Arts, Education, Commerce and Agriculture of The Ohio State University completed a four-year undergraduate program in three years or less, from 1941 through 1945. (*4*) The average academic record of these one hundred and four accelerates was found to be better than the records of one hundred and four other women graduating in the same period who entered at the same age, made approximately the same scores on the test of general ability given at entrance, and took the same type of program, but who spent the usual four years (three years and nine months, or four years if graduation was in August) in obtaining an undergraduate degree. The accelerates also participated reasonably in extra-curricular activities. And a follow-up a couple of years after graduation showed them beginning their careers in as promising fashion as the paired 'regular' cases.

Terman's notable continuing studies of his gifted child group have added impressive new evidence of the value of rapid progress in school of able young people. (*5, 6*) The Ford Foundation projects have reëmphasized the values of acceleration at the college level! Under all these circumstances, a further follow-up of the writers' war-time accelerate cases seemed desirable, to see how they were doing approximately ten years after graduation, and how they then

regarded their accelerated programs. This paper is a brief summary of such a study.

## CASES AND METHOD

When the above mentioned one hundred and four women accelerates were first being studied, it was hoped that a follow-up might be made. Hence their names, as well as the names of students paired with them but taking the usual three years nine months or four years to complete an undergraduate program, were saved. In 1946, forty-one more young women completed an undergraduate degree in three years or less, they were 'paired' with forty-one other women students. The present study is concerned with the total one hundred and forty-five accelerates and their paired cases who graduated from 1941 through 1946 and is based upon data gathered in the spring of 1954. The subjects at that time averaged around thirty-two years of age and had graduated an average of ten years previously.

For sending to the total of two hundred and ninety alumnae, a simple and straightforward two-page questionnaire was prepared, which could be very quickly filled in, almost entirely by checking. It was mimeographed on University stationery and began with the following statement:

> Dear Ohio State University Alumna:
>
> You were a student at O.S.U. during the Second World War, when certain adjustments of program were made. Campus life was in many respects different from normal times, and some students worked in war industry. An appraisal of such war-time adjustments and conditions as seen in retrospect seems desirable, especially since the cold war and possible crises might again bring educational changes. Your answers to the following questions can be very helpful; most require only a check after the answers you think correct.

Then followed in order questions about employment while in school, acceleration, extra-curricular activities, further schooling after the undergraduate degree, marital status, employment, and participation in community affairs. Special effort was made to phrase the entire document so that the recipients would not know that acceleration was a special matter of concern. And indeed the aim was, as the introductory statement said, to obtain broad appraisals regarding student life at the University in the war period.

UNIVERSITY OF MICHIGAN LIBRARIES

One of the total two hundred and ninety students was found to have died, and persisting search failed to locate the addresses of ten others. To the remaining one hundred and thirty-nine accelerates, and one hundred and forty students who took the regular time, the inquiry forms were mailed. Since the post office did not return them as undelivered it is assumed that the addressees received them, though a few may have very likely been accepted by a member of the family, or otherwise not come to the attention of the addressees. With each inquiry form was included a stamped addressed envelope for return.

The forms were mailed out April 7, 1954, and a follow-up was sent to those from whom replies had not then been received, on June 11. By August 1, eighty-one per cent (one hundred and twelve cases) of the accelerates had returned filled-in forms, and so had seventy-one per cent or one hundred cases of the regulars. A greater return from the accelerates might be attributed to greater interest on their part because they were aware of having adjusted more to the war-time situation. A further finding can perhaps be explained similarly. At the end of the questionnaire was the statement: "If you would like a report of the results of this study, give your name and address. But if you prefer, return the form unsigned." Eighty per cent of the accelerates, but only sixty-eight per cent of the regulars gave their names and addresses.

Though seventy-six per cent return on a questionnaire (for the two groups together) might be considered good, the question is as always as to whether selective factors may seriously have biased the results. Replies ranged from the highly favorable to the strongly and anonymously critical; it is believed that they are reasonably representative of the total group solicited. And the great majority of replies seem to have been carefully considered; additional written commentaries often extended to the backs of the blanks. In general, these alumnae seemed pleased to have had their opinions asked, and were interested in the topics.

## RESULTS

The first question asked was "During the period when you were an undergraduate, were you ever employed? If so, indicate when. While in school? —" and so on. The Table I summarizes the answers. It shows that the great majority found it possible to earn some money during the years when they were going to college.

TABLE 1.—PERCENTAGES OF 112 ACCELERATES AND 100 REGULARS WHO, DURING THE PERIOD WHEN THEY WERE UNDERGRADUATES, WERE EVER EMPLOYED

| | Acc. | Reg. |
|---|---|---|
| Were at some time employed | 79 | 86 |
| employed while in school | 73 | 68 |
| during vacations or quarters not in school | 31 | 65 |
| full time | 12 | 30 |
| part time | 69 | 73 |
| Percentages of those employed who felt that employment | | |
| put a strain on health | 3 | 7 |
| interfered with schooling | 7 | 8 |
| unfortunately limited social life | 19 | 18 |
| was a good experience | 93 | 87 |

That slightly fewer of the accelerates did so might be the result of not only more unremitting school-going or heavier loads but also less need, since completing a degree in three years obviously costs less than taking a year more. And, since accelerates and their controls began school at the same age, the accelerates finished a year younger.

Yet more interesting are the last four lines of the table. Only rarely did either group consider that employment put a strain on health, and slightly fewer of the accelerates than of the regulars so thought. Neither did either group feel that employment interfered with schooling. Only a few thought that it unfortunately limited social life, with no significant difference between the two groups. The great majority—and slightly more of the accelerates—thought that employment during the school years was a good experience.

The first part of Table 2 shows the percentages of the accelerate and the regular groups who ever took a four-quarter schedule, obtained credit by examination, took extra heavy course loads, or tried any combination of the above three methods of expediting progress through a college program. A few of the students who took four years for a degree did at some time take one of these steps; but they then were out of school, were ill, or for some reason did not actually shorten their over-all time in college. On the other hand, ninety-two per cent of the accelerate group, at least once, went four quarters in school; over half took extra heavy schedules; twenty-nine per cent got some credit by examination; and over

UNIVERSITY OF MICHIGAN LIBRARIES

TABLE 2.—EXPERIENCE WITH, AND JUDGMENTS OF, 'ACCELERATION'

| | Acc. | Reg. |
|---|---|---|
| Percentages of each group who ever | | |
| took a four-quarter yearly schedule | 92 | 11 |
| obtained credit by examination | 29 | 12 |
| took extra-heavy course schedules | 57 | 15 |
| used two or more of above methods | 57 | 6 |
| Percentages who felt that accelerated programs | | |
| were a strain on health* | 7 | 15 |
| prevented best school work* | 12 | 15 |
| undesirably limited social life* | 9 | 16 |
| were a desirable challenge | 62 | 40 |
| desirably saved time | 79 | 34 |

* Included here are a few judgments that undesirable results occurred at times though not generally, and a few replies of doubtful or 'not sure'.

half used two or more of such methods. In the judgment of the writers, there should be more use of credit by examination; and four-quarter schedules should then become less necessary. (*4, 7*)

Even though there may have been too continuous school-going and too much heavy course work, the second part of the table indicates that very few of the accelerate group thought accelerated programs a strain on health, not many judged that acceleration "prevented the student from doing as good work as he normally would," or that accelerated programs "undesirably limited social life." Instead, those not accelerating thought, more frequently, that such undesirable outcomes followed. And two-thirds of the accelerate group thought that acceleration presented a desirable challenge, while over three-fourths felt that desirable time was saved. Apparently few of the accelerates looked back on their college programs some ten years or so earlier with any feeling that damage had so been done them; rather, most felt that their acceleration had brought gains.

It has been held that accelerated programs cause too unremitting work and interfere with student activities, these being evaluated as making major contributions to a student's real growth in the college period. The writers' studies during and immediately after the war showed the accelerates only slightly less active than students taking the regular four years for an undergraduate degree. But it was felt desirable to get the reaction of these alumnae on

TABLE 3.—PERCENTAGES IN EACH GROUP WHO ENGAGED IN ACTIVITIES

| | Acc. | Reg. |
|---|---|---|
| Activities related to War, as Red Cross | 65 | 65 |
| three or more such activities | 9 | 11 |
| Usual student activities | 76 | 88 |
| three or more such activities | 42 | 60 |
| Percentages listing two or more who felt that activities | | |
| took too much time and energy | 10 | 15 |
| were sometimes too frivolous | 18 | 18 |
| were of value in broadening experience | 86 | 87 |

these matters, as they saw them ten years after. Table 3 summarizes the results. About as many of the accelerates engaged in activities related to the war as did the regulars. But more of the regular students went into the usual student activities—doubtless because they had more time, and partly probably because student activities are organized in terms of the conventional four-year autumn-winter-spring schedule. The last three lines of the table show a few more of the regular students feeling that these activities took too much time and energy and slightly more of the accelerates feeling that they were too frivolous, both groups agreeing that they were of value.

In considering Table 4 it is important to keep in mind that the accelerates and their paired regulars have been out of school the same length of time but that the accelerates were approximately a year younger, since accelerates and controls were selected as graduating in the same year and entering at the same age but the accelerates had been in school one year less. Thus in the same time since the undergraduate degree and though a year younger, twice as many of the accelerates have earned a further degree. In fact, the accelerates have earned thirty-two degrees: twenty-four Master's, one L.L.B.; one M.D.; four Ph.D.'s; and two Physical and Occupational Therapy certificates. In contrast, the only further degrees the regulars obtained were twelve Master's.

Slightly more of the regulars are married, slightly more of those married have children; slightly more of the accelerates have been divorced or separated. The accelerates married about a year younger; being as a group a year younger, they have been married about the same time. Whether such slight differences in the mar-

UNIVERSITY OF MICHIGAN LIBRARIES

TABLE 4.—FURTHER SCHOOLING, MARITAL STATUS, EMPLOYMENT, AND COMMUNITY ACTIVITIES

| | Acc. | Reg. |
|---|---|---|
| Percentages who have had further schooling | 51 | 43 |
| Earned further degrees | 24 | 12 |
| Now married | 80 | 85 |
| divorced or separated | 5 | 2 |
| median age of marriage | 22.3 | 23.2 |
| percentage of married having one or more children | 83 | 90 |
| single | 14 | 13 |
| Occupational status: housewife | 71 | 83 |
| employed (some also housewives)* | 39 | 28 |
| In at least one community activity (as church) | 84 | 84 |
| In three or more | 44 | 49 |

* Of the married accelerates, 29% were employed; of the married regulars, 16%.

riage, divorce and child-bearing rate are significant, especially in view of the differences in age, can hardly be judged. Slightly more of the regulars are housewives, and appreciably more of the accelerates are employed—substantially more of the married accelerates are also working. The numbers in community activities are approximately the same. In short, the two big differentiae are the substantially greater number of accelerates obtaining advanced degrees and substantially greater number continuing some sort of a career after marriage. The accelerates included a physician, an attorney, three chemists, a bacteriologist, etc.

As mentioned earlier, there were in various places opportunities for the respondents to comment, and a special invitation so to do at the end of the blank. The comments were various. Some accelerates felt that the program, though desirable, was so only in an emergency. Others were delighted at the saving of time, and considered that such possibilities of time-saving should always be available. Some complained of summer attendance, but others liked the contact then with more mature students. Accelerates who had been out in full-time employment tended to like the business-likeness of a four-quarter schedule; one remarked that simply the four ten-day periods between quarters gave her more vacation than she had had on the job. Some complained of heavy work, but others liked it, and said they worked better under some pressure.

There were sundry comments about curricula and instructional methods. Several alumnae wished that, somewhere, they had been taught to type; a number of the married women wished they had had courses in homemaking and child care. Some of these young women felt that their curricula should have prepared them more adequately for their vocation. Others wished that their programs had been broader, and, in fact, a number of the regulars had 'accelerated' a bit (for instance, gone to a summer session) to enrich their schedules with extra electives; three had obtained two undergraduate degrees in four years. It was pleasant to find, in this time of widespread criticism of schools and colleges, that many of the comments were favorable; most alumnae looked back upon their college years as pleasant and profitable.

## GENERAL SIGNIFICANCE OF THE FINDINGS

The present findings regarding acceleration, as viewed by students ten years after graduation, seem in substantial accord with results obtained earlier, and by other investigators (*1*, *3*, *4*, *5*). It seems clear that many students of good ability can complete the usual four-year undergraduate program in less time without any unfortunate outcomes. Several in the present study so finished in two calendar years. Further, most of the accelerates in this study (and most others) chose accelerated programs and carried them forward without any special counseling in the matter; a few should have been warned away from acceleration, and a majority might have planned better. The writers have found that with careful selection and guidance, outcomes are even better, as has the Fund for the Advancement of Education (*1*, *4*). Especially does it seem desirable that more use be made of credit by examination, for secondary school work beyond entrance requirement, also sometimes for crediting independent study (*4*, *7*). But the question now is as to how such findings are to be explained.

The accelerates in the present study were not notably superior in general ability to the average graduating student—they had tested only about ten percentiles higher on the test of general ability given at entrance to college. The possibility comes to mind that perhaps many undergraduate programs are sufficiently easy, studded with vacations, sometimes repetitious that some time-saving could become more general, or the programs be made more substantial.

UNIVERSITY OF MICHIGAN LIBRARIES

But the accelerates in important respects actually did better than the controls: in academic record, number obtaining advanced training, and number continuing a career. If they had been men, presumably yet more of them would have thus pushed ahead. Previous research of the writers, and the findings of Terman regarding his gifted group, indicate that accelerates are more likely to be successful and even distinguished in their careers (*4*, *5*). How are these anomalous findings to be explained: better schooling in less time, and better career too?

In most studies above referred to, the accelerates were equated with controls in initial ability, but not in ambition and energy. The person with more drive may more often accelerate, and work harder in school and job. Certain other studies have been broad enough or so planned that they somewhat allow for these factors, and still the accelerates did better. Might acceleration in itself have beneficial effects?

The accelerates not infrequently mentioned the challenge of such an intensive program, and the stimulus to confidence and ambition resulting from success in it. An opportunity into desired advanced training and career a year sooner can be very appealing to able young people. Earlier graduation may obviously leave more time and funds for further training. Certain other possible factors seem worth mention. May it perhaps be that in the growth years the organism is in subtle respects more ready for the broad learnings of undergraduate curricula, that as it becomes mature it is constitutionally set for adult career, and that careers begun early enough to be fully energized by the vigor of young adulthood are most likely to be successful? These last are indeed hypotheses. But there seems much to suggest that they have some soundness (*2*, *4*, *5*, *6*).

## SUMMARY

This paper reports a follow-up, some ten years after graduation from college, of a group of young women who during the war years completed a four-year undergraduate program in three years or less, and of cases then paired with them as to year of graduation, ability and age at entrance, and nature of college program. A previous study had shown the accelerates doing slightly better academic work than their controls and participating about as much in campus activities. The question was as to after-graduation careers of the two groups, and their retrospective judgments about their college programs.

A brief inquiry form was mailed to all cases (one hundred and thirty-nine accelerates and one hundred and forty regulars) for whom addresses could be found. Returns were obtained from one hundred and twelve accelerates and one hundred regulars (eighty-one per cent of the first and seventy-one per cent of the second group).

The one-time accelerates reported almost as much employment while in school; and few accelerates or regulars felt that such employment unfortunately affected health, schooling, or social life. Most thought such work experiences of value.

Most common method of acceleration was a four-quarter schedule, but a considerable number of students also obtained credit by examination, took heavy course loads, or used some combination of the above three methods. Very few accelerates thought their accelerated programs unfortunately affected health, school work, or social life. A majority felt acceleration a desirable challenge and saving of time. The non-accelerates were throughout more critical and dubious about acceleration.

About as many accelerates as non-accelerates took part in extra-curricular activities related to the war, but somewhat fewer participated in the usual undergraduate activities. Since there are few such in the summer, and most are organized on the presumption of a four-year program, the accelerates might be considered participating relatively well.

Slightly fewer of the accelerates have so far married and had children, but it must be kept in mind that they are a year younger than the regulars. The accelerates married almost a year younger, or about the same time after graduation. About the same number in each group are now active in community affairs. Twice as many of the accelerates have earned further academic degrees, as compared to the non-accelerates. Almost twice as many of the married accelerates are continuing a career.

The above findings are explained as possibly in part due to a greater ambitiousness and energy on the part of the accelerates, but probably also a result of certain challenges and advantages of accelerated programs. The considerations involved seem of fundamental importance in planning secondary and higher education, and more so for men and in furtherance of career than for women as in this study.

UNIVERSITY OF MICHIGAN LIBRARIES

REFERENCES

1) *Bridging the Gap Between School and College*. New York: The Fund for the Advancement of Education, 1953.

2) H. C. Lehman. *Age and Achievement*. Princeton: Princeton University Press, 1953.

3) G. Melcher. *An Evaluation of the Northeast Accelerated Junior College*, Kansas City, Mo.: Board of Education, 1953.

4) S. L. Pressey. *Educational Acceleration: Appraisals and Basic Problems*, Columbus, Ohio; Ohio State University Press, 1949.

5) L. M. Terman and M. H. Oden, *The Gifted Child Grows Up*. Stanford, Calif.: Stanford University Press, 1947.

6) L. M. Terman. "The discovery and encouragement of exceptional talent," *American Psychologist*, 1954, 9, 221–230.

7) Mazie E. Wagner. *Anticipatory Examinations for College Credit: Twenty Years Experience*. Buffalo: University of Buffalo Studies, Vol. 20, No. 3, December, 1952.

# A NOTE ON THE STANDARD ERROR OF MEASUREMENT

C. H. PATTERSON

Veterans Administration Regional Office

St. Paul, Minnesota

The standard error of measurement of a test score has sometimes been suggested as being more useful in counseling the individual than other measures of reliability. There appears to be some disagreement on the meaning of this statistic, and thus of the nature of the statements which are made concerning it. This problem was aired in the exchange of papers between Guilford and Garrett in 1937 (*1*, *2*, *4*), in which no apparent agreement was reached.

Two main types of difficulty appear in interpreting the standard error of measurement. The first concerns the nature of the statements which can be made about the true score of an individual with a given obtained score, in terms of the standard error of measurement. The second difficulty, more general in nature, concerns the type of statement which can be made in any situation where probabilities are involved.

The first problem basically involves the meaning to be assigned to an obtained score on a test, and its relation to the 'true' score corresponding to a given obtained score. Some writers, e.g., Guilford (*3*) and Lindquist (*9*) state that no exact statement can be made concerning the true score of an individual making a given obtained score. They insist that the standard error of measurement applies to the observed score, and not to the true score. Since the standard error of measurement is the standard deviation of the distribution of the scores which an individual would make on repeated tests, statements must be limited to these obtained scores. The mean of this distribution of obtained scores would be the individual's true score. Thus, a typical statement would be: "If a certain individual has a true score of 50 points in the test, then we may expect two-thirds of all his actual scores to lie between 46 and 54" (*3*, p. 414), given a standard error of measurement of 4. Lindquist (*9*, pp. 221–222) follows a similar interpretation. Gulliksen also agrees that "no probability statements can be made to



UNIVERSITY OF MICHIGAN LIBRARIES

apply to all persons who make a given *observed* score," but only for all persons with a given true score.

These writers suggest that the only way in which statements about true scores may be made is by means of the standard error of estimate of a true score predicted from an obtained score by means of a regression equation. This method is seldom used because of the labor involved.

Another group of writers make statements concerning the true score on the basis of an obtained score and its standard error (e.g., *7*, *10*). They do this by taking the obtained score as the estimate of the true score, either explicitly or implicitly, and using the standard error of the observed score as being equivalent to the standard error of a true score. Kelley writes: "$\sigma_{1_\infty}$ may correctly be used to indicate the standard error when $x_1$ (an obtained score) is estimated from a knowledge of $x_\infty$ (a true score) (a process actually never performed) or when $x_\infty$ is estimated by using non-regressed x." (*8*, p. 403). This method leads to interpretations such as the following (given a score of 70 and a standard error of 7.72): "There is one chance in six that John's true ability lies below 62.3, so that there are four chances in six that it lies between 62.3 and 77.7" (*7*, p. 172).

Some writers make statements of the type that the true score of an individual with a given obtained score is 'almost certain' to lie within three standard errors of the obtained score. Peatman, for example, referring to confidence limits of an IQ of 109, given a standard error of 4.5, states: "thus it is likely that persons with an IQ score of 109 have parameter (i.e., true) scores whose values lie between 100.0 and 118.0; and we can be quite confident that the parameter values will not be less than 95.5 or greater than 122.5" (*11*, p. 379). These statements do not seem to be exact probability statements, though they may frequently be taken as such by readers. They thus do not appear to be of the type which implicitly accept an obtained score as an estimate of the true score. They may be of a type described by Gulliksen (*6*, p. 20) in which 'reasonable limits' are assigned to the true score.

While the observed score is an estimate of the true score, it is not the true score, nor the best estimate of the true score. It would thus appear that statements based on the acceptance of the obtained score as equivalent to the true score are questionable. Statements concerning the true scores of individuals with a given observed score should strictly be based on the regressed score and its standard error of estimate. While the effort involved in this

procedure of estimating true scores may not be justified, statements concerning the true scores associated with given observed scores should probably be limited to regressed scores. Rigorous logic would indicate, therefore, that statements should be of the type used by Lindquist, Guilford and Gulliksen.

Nevertheless, it appears that the use of the obtained score as evidence of the true score is common. The correct form of statement when this is the case is of the form that for a given observed score the corresponding true score will fall within a given interval, determined by the observed score, its standard error, and the confidence level desired, and we shall be right in this statement in a specified proportion of cases in the long run corresponding to the confidence coefficient selected.

This leads to the nature of probability statements, regarding confidence intervals within which a given parameter, such as a true score, lies. The true score is a fixed value, and not a random variable. Therefore, in a particular case it either is or is not within a given confidence interval, i.e., the probability is either unity or zero, respectively. Probability should be attached to the statement, which is true in a given proportion of the cases in the long run, and not to a score. It cannot be stated, therefore, that the probability is two-thirds, or 68 out of 100, that a given individual's true score lies within plus or minus one standard error of his observed score taken as an estimate of his true score. The correct statement is that a given individual's true score lies within plus and minus one standard error of his observed score, and we shall be correct in this statement for two-thirds of the cases, or approximately 68 times out of 100.

Following the frequency theory of probability, statements must be of this type, the probability applying to the truth or falsity of the statement in a series of cases, not to a score. Following the interpretation of Guilford, Lindquist and Gulliksen, the correct statement would be that a given individual's observed score lies within plus and minus one standard error of his true score, and we shall be correct in this statement in approximately 68 cases out of 100. The correct form of statement regarding a true score estimated by means of the regression formula would be that for all cases with the same estimated true score, the actual true score falls within three standard errors of the estimated score over ninety-nine per cent of the time.

It is possible to make certain statements about the scores of

UNIVERSITY OF MICHIGAN LIBRARIES

groups of individuals. Using the concept of tolerance limits, it appears to be possible to assign exact probabilities to statements referring to a group of individuals with a given obtained score.[1] Usually, however, it is sufficient to be able to state that in repeated testing the obtained scores of a group of individuals with a given obtained score on the first test will almost certainly fall within three standard errors ('reasonable limits') of the corresponding true score. It is possible to state that "for all persons with a given true score, the probability is over .997 that the observed score will lie between plus or minus three times the standard error of measurement for that true score" (*6*, p. 20).

While there may seem to be little practical significance to the differences in statements discussed above, it is important that those using the standard error of measurement understand the nature and the bases of the different types of statements. It is hoped that this brief discussion will serve to clarify these differences, which, to the knowledge of the writer, have not been clearly differentiated heretofore.

REFERENCES

1) H. E. Garrett, "On the interpretation of the standard error of measurement." *Amer. J. Psychol.*, 1937, 49, 679–680.

2) — — — "A rejoinder." *Amer. J. Psychol.*, 1937, 49, 683–685.

3) J. P. Guilford. *Psychometric Methods*. New York: McGraw-Hill, 1936.

4) — — — "More concerning the interpretation of the standard error of measurement." *Amer. J. Psychol.*, 1937, 49, 680–683.

5) — — — *Fundamental Statistics in Psychology and Education*. (2nd Ed.) New York: McGraw-Hill, 1950.

6) H. Gulliksen. *Theory of Mental Tests*. New York: Wiley, 1950.

7) T. L. Kelley. *Interpretation of Educational Measurements*. Yonkers, N. Y.: World Book Co., 1927.

8) — — —. *Fundamentals of Statistics*. Cambridge: Harvard University Press, 1947.

9) E. F. Lindquist. *A First Course in Statistics*. (Rev. Ed.) Boston: Houghton Mifflin, 1942.

10) Q. McNemar. *Psychological Statistics*. New York: Wiley, 1949.

11) J. G. Peatman. *Descriptive and Sampling Statistics*. New York: Harper, 1947.

---

[1] For the method of doing this, see "The interpretation of the standard error of measurement," by the author in the March, 1955, issue of the *Journal of Experimental Education*.

# THE STABILITY OF INTEREST SCORES OF COLLEGE FRESHMEN

CLAY E. GEORGE and ALBERT J. KINGSTON

The Agricultural and Mechanical College of Texas

It is a common practice among colleges and universities to administer some type of psychological tests to entering freshmen. Many of these instruments are designed to evaluate a student's aptitude for academic work or to appraise his achievement in certain subject matter areas. The results of these tests often are employed for placement purposes. Some institutions also administer an interest inventory at the same time, the results of which often are utilized during subsequent counseling throughout the student's freshman year. The research of Strong (*3*) has indicated that the interest patterns of young men of college age tend to fluctuate considerably. As it is axiomatic in counseling that a counselor must have confidence in any test results which he employs, it was felt that a study of the stability of the interest scores of college men might be valuable.

Three approaches were decided upon. First, the test re-test means were compared for each of the ten scores of the *Occupational Interest Inventory* (O. I. I.) (*1*). Next, the profiles (pattern) of the scores were tested by computing rho's between the test re-test scores for a large number of students. The final approach was to make a frequency distribution of the number of students scoring higher or lower on the nine fields and types of interest to various degrees.

## METHODOLOGY

On September 5, 1953, all entering freshmen at the Agricultural and Mechanical College of Texas were administered the *Occupational Interest Inventory*. On October 27, 1953, a randomly selected group of two hundred fifty-six male students were readministered the inventory.

The major purpose of the study was to determine the stability of individual patterns rather than group patterns. In order to obtain this information three different techniques of analysis were employed. First, the means and standard deviations for each field



UNIVERSITY OF MICHIGAN LIBRARIES

TABLE I.—TEST RE-TEST MEANS AND STANDARD DEVIATIONS FOR FIELDS, TYPES, AND LEVELS OF INTEREST

| Scale | Pers. Social | Fields of Interest | | | | | Types of Interest | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Nat. | Mech. | Bus. | Arts | Sci. | Verbal | Manip. | Comp. | Level of Interest |
| Test Mean | 14.86 | 22.10 | 22.46 | 19.00 | 13.45 | 25.77 | 7.62 | 12.45 | 9.81 | 69.95 |
| SD | 5.12 | 6.50 | 6.29 | 8.15 | 9.54 | 6.41 | 4.44 | 2.49 | 3.95 | 10.12 |
| Re-Test Mean | 14.96 | 22.33 | 22.19 | 20.63 | 15.99 | 24.59 | 7.73 | 11.75 | 9.45 | 69.76 |
| SD | 6.68 | 9.79 | 5.48 | 8.14 | 7.74 | 7.52 | 4.73 | 3.71 | 4.31 | 8.75 |
| D | .10 | .23 | −.27 | 1.63 | 2.54* | −1.18 | .11 | −.70* | −.36 | −.19 |

* P = .05

and type of interest, as well as the level of interest, were determined and compared. Table I illustrates the relationships obtained by this comparison.

This analysis seemed to indicate that there was little difference between the average scores of the group on the initial test and on the re-test. Only two scores seemed to be significantly different—the Arts score, showing a slight gain, and the Manipulation score, showing a slight loss. Both were significant at the 5 per cent level of confidence.

In order to determine whether there were significant changes among the various factors which comprised a student's pattern of interests, the scores of one hundred selected men were analyzed. Rho coefficients were calcuated between the percentile scores of each field of interest, type of interest, and level of interest obtained

TABLE II.—RANK-DIFFERENCE CORRELATIONS BETWEEN PERCENTILE SCORES ON THE TEN INTEREST SCALES OF THE O.I.I. FOR 100 RANDOMLY SELECTED MALE FRESHMEN

| No. of Students | Rho | P |
|---|---|---|
| 3 | −.20– .00 | N.S. |
| 17 | .01– .54 | N.S. |
| 11 | .55– .64 | <.10 |
| 14 | .65– .74 | <.05 |
| 55 | .75–1.00 | <.01 |

TABLE III.—CHANGES IN PERCENTILE RANKINGS OBTAINED BETWEEN FIRST AND SECOND TEST (FIELDS AND TYPES OF INTEREST)

| | No Change in Score | Percentile Score Increase | | | | | Percentile Score Decrease | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1–19 | 20–39 | 40–59 | 60–79 | 80+ | 1–19 | 20–39 | 40–59 | 60–79 | 80+ |
| Pers.-Soc. | 68 | 32 | 30 | 7 | 1 | 2 | 49 | 57 | 8 | 1 | 1 |
| Nat. | 87 | 58 | 29 | 8 | 0 | 0 | 52 | 19 | 1 | 2 | 0 |
| Mech. | 73 | 55 | 34 | 4 | 0 | 0 | 39 | 46 | 5 | 0 | 0 |
| Bus. | 76 | 51 | 34 | 7 | 1 | 0 | 41 | 39 | 6 | 1 | 0 |
| Arts | 63 | 57 | 47 | 10 | 0 | 0 | 49 | 26 | 4 | 0 | 0 |
| Sci. | 66 | 46 | 24 | 4 | 0 | 0 | 47 | 44 | 20 | 4 | 1 |
| Totals | 433 | 299 | 198 | 40 | 2 | 2 | 277 | 231 | 44 | 8 | 2 |
| Verb. | 72 | 36 | 44 | 8 | 3 | 0 | 44 | 36 | 13 | 0 | 0 |
| Manip. | 50 | 40 | 53 | 13 | 2 | 0 | 42 | 41 | 13 | 1 | 1 |
| Comp. | 70 | 38 | 29 | 6 | 1 | 1 | 46 | 52 | 11 | 2 | 0 |
| Total | 192 | 114 | 126 | 27 | 6 | 1 | 132 | 129 | 37 | 3 | 1 |

on the initial test and the re-test. The rho's for each of these one hundred students are given in Table II. The results of this analysis indicated that the general pattern of interest scores had undergone considerable change in at least twenty cases.

This analysis would seem to indicate that the interest patterns of at least twenty of the one hundred students had fluctuated so widely that use of the initial test would prove to be very misleading.

One further technique of analyzing the data was employed. It was felt that analysis of the changes in each interest scale according to the percentile rankings provided by the authors of the inventory would prove helpful in illustrating the degree of fluctuation found in some individuals. Table III illustrates these findings.

## CONCLUSION

This study may be regarded as substantiating the axiom that test results must be interpreted with caution. The measured in-

UNIVERSITY OF MICHIGAN LIBRARIES

terest patterns of certain male freshmen had changed radically after a period of less than two months of college work. It seems likely that a counselor who relied on the results obtained at the initial inventory would err in interpreting the interests of many counselees.

This study also illustrates the error one may make by relying solely upon test re-test averages in studying score stability. From the averages given in Table I, it would seem that there was very little change in either pattern shape or subtest levels. Tables II and III indicate, however, that there were a very large number of marked individual changes in both profile shape and level.

Reid (*2*) conducted a similar study with the *Kuder Preference Record*. He reports rank order coefficients ranging from −.72 to 1.00, with a median of .88. These were based on freshman college student interest patterns as measured at entrance and fifteen months later. In the present study the range of rhos was −.20 to 1.00 with a median of .77. Strong (*4*) reports a correlation of .81 between permanence of interest scores and time interval between test and re-test, with the influence of age partialed out. Since the Kuder tends to be more stable over a fifteen-month period than the Lee-Thorpe over a two-and one-half-month period, and permanence decreases with time, it would appear that the Kuder is the more stable instrument. Sampling differences may have affected these results, however, so this conclusion is only tentative.

## REFERENCES

1) F. A. Lee, and L. P. Thorpe. *Occupational Interest Inventory*. Advanced Series. Los Angeles: California Test Bureau, 1943.

2) J. W. Reid. "Stability of measured Kuder interests in young adults." *J. educ. Res.*, 1951, 45, 307–312.

3) E. K. Strong. *Vocational Interests of Men and Women*. Stanford: Stanford University Press, 1943.

4) — — — "Permanence of interest scores over twenty-two years." *J. appl. Psychol.*, 1951, 35, 89–91.

# PRELIMINARY EXPLORATION OF PERSONALITY FACTORS RELATED TO SUCCESS IN A HIGH-LEVEL MILITARY SCHOOL

JAMES H. L. ROACH and EDWARD M. WALL

The Air University

In the Air Command and Staff School, as in other educational institutions, we have found that intellect and achievement are far from perfectly correlated. One of our best predictors of success in this course (r = .64) has been the Initial General Survey Test (IGST) which was constructed to measure military and general background knowledge.

However, in the AC&SS class '53-B there were fourteen officers who fell in the 'C' group on the IGST but whose achievement in the course was out of line with expectancy. Of these fourteen, six ended in the 'A' group academically (top seven per cent) and eight in the 'E' group (bottom seven per cent).[1] When it was further noted that the ACE (American Council on Education Psychological Examination) scores of these two groups were essentially the same, their differences in achievement became even more interesting. An attempt was then made to find if there were any measurable personality differences between the two groups which could account for the differential.

Faculty advisors of these students, who had worked with them closely, were asked to rate their students on a specially devised scale listing twenty variables. Scores on each item ranged from zero to five. Advisors were not told why these students were selected for this rating.

Moreover, during the course, fellow students had filled out officer effectiveness report forms on each other as a part of their regular course program.[2] From these reports about twenty-five independent

[1] Final standing in the course was based upon four tests and a final examination, two seminar grades, writing and speaking grades and a staff study report. Large, significant differences between the two groups of subjects were found on all of the above factors except the staff study report.

[2] Student officers were required to complete these forms on each fellow



UNIVERSITY OF MICHIGAN LIBRARIES

evaluations on six variables, (five of which appeared on the faculty's scale), were available for each of the fourteen subjects. Scores on this scale ranged from one to sixteen. (A five-point scale each broken into three subdivisions with one subdivision at the bottom end of the scale.)

The faculty rated the 'over-achievers' higher on each of the twenty factors; and for ten of these the differences were statistically significant (Table 1). The ten significant factors with general description were as follows:

1) Personality: closed, withdrawn or open, pleasant.

2) Emotional Stability: tense, upset, involved, or free from 'blowups'.

3) Leadership: ability to command, exert control, inspire confidence; skill in directing others.

4) Activity level: degree of participation in general activities.

5) Reliability and responsibility: degree to which accepts responsibility and can be relied upon to execute assigned tasks.

6) Personal adjustment: problems faced and ease in solving them.

7) Judgment: ability to grasp a situation, soundness and logic of thought.

8) Initiative: frequency of initiating and explaining projects and ideas.

9) Originality: creativeness and originality of thinking.

10) Sports activities: degree and quality of participation.

Factors in which the differences were not statistically significant were: persistence, general health and physical well-being, family adjustment, job knowledge, coöperation, growth potential, motivation, tact, study habits, and social activities. Oddly, faculty members reported that they had more difficulty in remembering the better students.

Evaluations by fellow students also gave the over-achievers higher ratings on all six of their variables, but here all differences were statistically significant (Table II). Student ratings on judgment, leadership, and reliability and personal responsibility agreed closely with faculty ratings. However, students rated the over-achievers higher than the faculty did on coöperation, (ability and

---

officer in their seminar to: (1) Learn how to complete the form properly; and (2) acquaint students with an appraisal of their effectiveness as seen by fellow students. They were not used in course grades.

TABLE I.—MEAN SCORES OF OVER-ACHIEVERS AND UNDER-ACHIEVERS ON PERSONALITY FACTORS WHERE SIGNIFICANT DIFFERENCES WERE FOUND IN FACULTY RATINGS

| Factor | Mean Score Over-achievers | Mean Score Under-achievers | Differences in means |
|---|---|---|---|
| Originality | 4.20 | 2.00 | 2.20 |
| Initiative | 4.18 | 2.15 | 2.03 |
| Judgment | 4.18 | 2.23 | 1.95 |
| Personality | 4.18 | 2.31 | 1.87 |
| Personal adjustment | 4.22 | 2.38 | 1.84 |
| Activity level | 4.27 | 2.62 | 1.65 |
| Emotional stability | 3.73 | 2.23 | 1.50 |
| Leadership | 3.64 | 2.23 | 1.41 |
| Reliability and Responsibility | 4.55 | 3.15 | 1.40 |
| Sports activities | 4.00 | 3.08 | .92 |

TABLE II.—MEAN SCORES OF OVER-ACHIEVERS AND UNDER-ACHIEVERS ON PERSONALITY FACTORS WITH SIGNIFICANT DIFFERENCES FOUND IN RATINGS OF FELLOW STUDENTS

| Factor | Mean Score Over-achievers | Mean Score Under-achievers | Differences in means |
|---|---|---|---|
| Leadership | 11.17 | 8.00 | 3.17 |
| Growth potential | 11.17 | 8.38 | 2.79 |
| Reliability and Responsibility | 11.33 | 8.75 | 2.58 |
| Over-all evaluation | 8.67 | 6.13 | 2.54 |
| Judgment | 11.33 | 8.88 | 2.45 |
| Coöperation | 11.33 | 9.75 | 1.58 |

willingness to work in harmony with others), and growth potential (capacity to handle jobs of increased scope and responsibility, self-improvement efforts, ability to learn rapidly). The one variable rated by students and not faculty was over-all effectiveness, e.g., a comparison with all officers in the same grade.

Results of this study show clearly that certain favorable personality characteristics were exhibited to a much greater degree by the over-achievers than by the under-achievers. Inasmuch as these two groups were about equal in intellectual capacity as well as background knowledge, it is felt that these personality differences, especially leadership, judgment, reliability and personal responsibility, account in part, at least, for the superior success of the over-achievers. These findings are in general agreement with those re-

UNIVERSITY OF MICHIGAN LIBRARIES

ported by Terman, Cox, and others in respect to personality factors related to life success of gifted individuals.

## REFERENCES

1) Catharine C. Cox. *The Early Mental Traits of Three Hundred Geniuses.* Vol. II of Genetic Studies of Genius, L. M. Terman, (Ed). Stanford: Stanford University Press, 1926.

2) L. M. Terman, and M. H. Oden. *The Gifted Child Grows Up.* Vol. IV of Genetic Studies of Genius. L. M. Terman, (Ed). Stanford: Stanford University Press, 1947.

# BOOK REVIEWS

STEPHEN M. COREY. *Action Research to Improve School Practices.* New York: Bureau of Publications, Teachers College, Columbia University, 1953, pp. 161. $3.00.

Dr. Corey's proposal for action research springs from his conviction that dramatic changes must be made in public school curricula and practice if youth are to be prepared for the demands of a changing world. He believes that the knowledge gained from research is most likely to be translated into action if teachers and school administrators are ego-involved in the research.

The quantity of space given to the value and processes of group decision may surprise the reader, but the author emphasizes the necessity for enlisting the coöperation of all who will be affected by the changes. Reports of some actual projects are given to illustrate the steps in formulating and testing hypotheses.

Some distinguishing characteristics of action research are stated: it is focused on practical problems facing the participants; it is carried on amid the complexities of the school without the customary controls of the laboratory experiment; its value is measured in terms of improvement in educational practice; research design is subject to modification during experimentation; generalizations from research are extended vertically instead of laterally. It goes further than casual inquiry because evidence is systematically recorded, analyzed and interpreted.

Participants, Corey says, must be willing to admit and explore flaws in the school system, recognize that not all hypotheses will be proven, develop skills in leadership and group membership, evaluate group meetings, try to understand the feelings of colleagues, and provide time, resources and adequate consultant services. The last-named essential condition for success, the reviewers feel, may be the most difficult to arrange.

Chapter 7 presents some statistical evidence to support the vertical extension of generalizations and describes some basic statistical operations which might be used in action research. The author believes that the teacher will learn to use these techniques when he becomes aware that they will help him solve problems and answer questions.



UNIVERSITY OF MICHIGAN LIBRARIES

The book is addressed principally to school administrators and supervisors whose leadership and support would significantly influence the success of action research. Those who want to test practices and attack persistent school problems will find some useful information. Some readers will probably comment that more attention is given to improving staff harmony than to planning research on instructional problems. Others will infer that the working relations of faculty members will be reflected in the classroom and that principles and techniques presented in this connection can be applied to investigations of the school offering. Some published studies of action research and group decision are summarized by the author, and eighty references are included in the bibliography.

MAURICE D. AND JEANNE A. WOOLF
*Kansas State College, Manhattan*

SIR CYRIL BURT. *The Causes and Treatment of Backwardness.* New York: Philosophical Library, 1953. pp. 128.

American students of the problem of mental deficiency are familiar with Dr. Burt's contribution in this area. Particularly, are they familiar with the book published fifteen years ago called *The Backward Child*, where the problems of contributions of heredity and environment were considered more sanely and comprehensively than some of the American volumes of that period. Burt considers the present smaller volume an abridgment of this larger one. In abbreviating, Burt has humor enough to quote Boswell that shortening a book was too often "like cutting off the head and the tail of a cow," to which Dr. Johnson was supposed to have replied, "At its best it should be like making the cow produce a calf." Burt hopes that the book is not a simple curtailment, but that it is in many ways a fresh book with a different purpose. The purpose is to serve the staff of the National Children's Home with knowledge and information they can use in dealing with something like 3,000 boys and girls, housed in forty branches.

The suggestion for the book grew out of the lecture that Burt made at one of the Annual Convocation Lectures. This he was asked to enlarge and rewrite in book form. Material of the book is presented in eight chapters dealing with historical factors, methods of investigation, environmental factors, personality factors—phys-

ical, intellectual and emotional—and some practical conclusions. The book contains a list of references and a separate index of authors and subjects.

The references used and all other points of reference are distinctly English-bound. This is understandable enough, since the setting and the history is also English. The aim of the book obviously is to search in psychology for knowledge that can be of help in dealing with backward children. One thing that Burt likes to make clear is that the problem is not one of 'physiology of the brain'. The ordinary man, says Burt, erroneously supposes that the expert on physiology is the physician—hence arises the widespread conclusion shared by a number of doctors that psychology is or ought to be a branch of medicine. The problem is not merely a dispute about words; the notion that educational subnormality is the result of a physical defect, which can be cured by glandular extracts, brain surgery, or a tonic for the nerves is widespread. Then Burt argues that the proper field, of course, is psychology, and that the first investigations were distinctly English. The investigator who first introduced the scientific approach to the psychology of the individual child was Sir Francis Galton. And it was Galton, Burt insists, who advocated that all such conclusions needed to be checked by new experimental and statistical techniques. To this end he invented an experimental device of mental testing and the statistical method known as correlation. The idea of mental testing, even in a child guidance clinic, according to Burt, was borrowed by Binet from Galton, and was introduced in America by Cattell, who was one of the first of Galton's fellow-workers. All of which is to prove that the kind of psychology that can help backward children originated in England and its most profound influences came out of England.

Nobody in the United States is going to argue against the importance of Galton or the fact that Cattell worked with him. But few, if any, would agree with the particular over-Anglicized emphasis by Burt about the origin of all things good in England that has to do with the study of backward children.

This bias does not detract from the sane and sensible, intelligent approach in considering the problems that Burt makes in this book as he has in his previous publications. For a straightforward presentation of the causes of, and methods of dealing with, backward-

UNIVERSITY OF MICHIGAN LIBRARIES

ness, this book can be of help to the teacher of backward children and students of the field. Stressed by Burt is what one would expect him to stress; namely, the need for records that should be objective so the analysis can be carried out according to a systematic plan, because, as he says in the last sentence in the book: "In this way, and this way alone, will the accumulated wisdom gained by each one be eventually placed at the disposal of all."

H. MELTZER

*Psychological Service Center, St. Louis*

BRUNO BETTELHEIM. *Symbolic Wounds, Puberty Rites and the Envious Male*. Glencoe, Ill.: The Free Press, 1954, pp. 286. $4.75.

Dr. Bruno Bettelheim, best known for his book about children, *Love Is Not Enough*, has made certain observations on the children in the Sonia Shankman Orthogenic School, University of Chicago, which prompted him to question certain psychoanalytic anthropological theories about circumcision and rituals. Out of these observations on neurotic and psychotic children has come this volume of what may be considered a re-interpretation of puberty rites and circumcision.

This, on Bettelheim's part, represents an attempt to separate fantasies of patients from historical reality. Contrary to the ordinary belief that initiation rites are imposed by society on the young as a means of introducing them into adult society, Bettelheim has come to almost the opposite conclusion; namely, that they are the efforts of children to help themselves and they are not prompted by elders or irrational drives. Instead, they are based on the urge for freedom from those drives and pressures to master the irrational forces. That's the thesis.

The evidence is based on Bettelheim's own observations and a second and a third look at some anthropological reports. These convinced him that there has been an over-emphasis on the assumed end and under-emphasis on the means of obtaining it; that there has been an over-emphasis on penis-envy in girls and castration anxiety in boys and the neglect of what might be considered the vagina-envy in boys—the fascination of boys with female breasts, lactation, pregnancy and child-bearing. This thesis is presented in

a series of chapters which, to some extent, represent arguing with himself about observations, theoretical considerations and interpretation of meanings of such material as Monotheism and the Castrating Father; Fertility, the Basic Cult; Castration, Circumcision and the Great Mother; Australian Rites; The Men-Women; The Secret of Men; and Girls' Rites. The author ends the book with an expressed wish that if we could satisfy through rituals, or through more civilized—less magic and more satisfying—institutions, the desires of both men and women to participate in the activities and enjoyments normally belonging to the other sex, each sex would gain greater inner autonomy, could better accept its own rôle and that of the other, that two could live with one another more satisfactorily. Instead of trying to solve these problems through ritual, he suggests we could be equally serious in searching for solutions that are more rational, more effective socially and more satisfying privately.

In some instances reported by Bettelheim, both from his own observations in school, based on verbal reports of children, and some observations on, say, the Australian aborigines, in whom he reports he found an absence of castration anxiety repressions, his emphasis is more justified than the over-emphasis given by psychoanalysts and anthropologists who have accepted psychoanalytic interpretations. Whether a whole book is needed to prove that the opposite theory is correct is a question. There is enough in here for easy reading, but to some extent the book is a stretched-out article. The references given are mostly from the anthropological field. Bettelheim is modest enough to recognize he is out of his field here.

H. Meltzer

*Psychological Research Center, St. Louis.*

Marvin D. Glock. *The Improvement of College Reading.* Boston: Houghton-Mifflin Company, 1954, pp. 307. $2.50.

The material presented in this text is derived from a college reading program developed and successfully used for several years at Cornell University. It is designed for use in any college reading program and in English courses which emphasize improvement of reading. Stress is placed upon improving comprehension, analytical skill, judgment, skimming ability, vocabulary study, and versa-

UNIVERSITY OF MICHIGAN LIBRARIES

tility in adjusting mode of reading to purpose and nature of the material. The importance of motivation is recognized throughout the course.

There are thirty exercises or lessons. In each exercise there are four parts such as skimming, reading to answer a problem, critical reading, and vocabulary development (Exercise 18). There are sets of questions on each part of each exercise. These questions are concerned with facts, understanding of ideas, tone and intent, organizational pattern, and vocabulary. Although speed is measured, comprehension is properly considered to be more important. A particularly good procedure is to check that reading is not too fast for comprehension and to assure that speed is adjusted to purpose and content. Diagnostic charts are provided for recording day-to-day improvement so that the student may check his own gains. Recognizing that a person cannot expect much improvement in vocabulary during a short-term program, emphasis is placed upon vocabulary exercises to acquaint the student with methods of vocabulary study such as the use of basic roots and affixes.

The discussions and procedures in this text are based upon sound experimental findings. The author has been successful in selecting and giving proper emphasis to the basic factors and skills necessary for developing proficient reading. This is one of the best programs of reading instruction at the college level encountered by the reviewer.

MILES A. TINKER

*University of Minnesota*

(*Concluded from Inside Front Cover*)

these charges, as well as information regarding the services of the American Documentation Institute in making available tables, etc., that for economic reasons may be omitted from a printed article, by writing to the Editors or the Publishers.

*Double-spacing.*—Manuscripts should be typed, written on one side of the paper only, and double-spaced throughout including quotations, footnotes, and bibliographical references.

*Footnotes.*—Footnotes are to be numbered consecutively beginning with '1', and should be on a separate sheet at the end of manuscript. (Footnotes to tables carry the *, †, and ‡.)

*Titles.*—Titles of articles should be brief, preferably three to eight words, with an extreme maximum of twelve words.

*Type style.*—Manuscripts are not to be marked for type style—this is done in the editorial office.

---

Books and other materials for review, and business communications should be addressed to THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Warwick & York, Publishers, 10 E. Center St., Baltimore 2, Md.

Subscribers should notify the Publishers of change of address at least four weeks in advance of publication of the issue with which change is to take effect; both the old and new address should be given.

The Publisher desires every subscriber to get all issues to which he is entitled. Each journal is securely enclosed in a sturdy wrapper on which the subscriber's name and address have been printed, and is delivered directly to the Post Office, postage prepaid. Second-class matter is handled less by postal employees than other mail; moreover, if the Post Office is unable to make delivery, a notice to this effect is sent the Publisher and the magazine returned. Consequently, it is doubtful if one journal in many thousands is actually lost in transit.

But after an issue has been delivered to the proper address many things may happen to it—it may be diverted, or misplaced, or borrowed and not returned. For this neither Post Office nor Publisher is responsible. However, a subscriber who does not find a given issue in its assigned place may innocently make a claim of non-receipt. No claim for non-receipt of an issue can be honored unless made within four weeks after arrival of the next succeeding number. In order that a claim may arrive within the time limit it should be addressed to the Publisher—not to an agency.

WARWICK AND YORK *Publishers* BALTIMORE 2, MD.

UNIVERSITY OF MICHIGAN LIBRARIES